

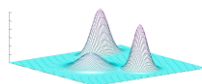


Book of Abstracts

II International Workshop on Proximity Data, Multivariate Analysis and Classification

October 20-21, 2016

Barcelona (Spain)



Copyright © 2016, of the texts and images: the authors.

Copyright © 2016, of the edition: the Working Group in Multivariate Analysis and Classification (AMyC) of the Spanish Society of Statistics and Operations Research (SEIO).

Editors:

José Fernando Vera, Universidad de Granada.

Eva Boj del Val, Universidad de Barcelona.

Edited by the Publications service of the University of Barcelona.

ISBN: 978-84-617-7320-6

Organizing Committee

José Fernando Vera, UGR (jfvera@ugr.es)

Eva Boj del Val, UB (evaboj@ub.edu)

Scientific Committee

Carles M^a Cuadras, UB (ccuadras@ub.edu)

Josep Fortiana, UB (fortiana@ub.edu)

José Fernando Vera, UGR (jfvera@ugr.es)

Albert Satorra, UPF (albert.satorra@upf.edu)

Eva Boj del Val, UB (evaboj@ub.edu)

M^a Teresa Costa, UB (tcosta@ub.edu)

Joan Guàrdia i Olmos, SCE (jguardia@ub.edu)

Important Dates

October, 14th: Deadline for abstract submission

October, 16th: Notification of acceptance

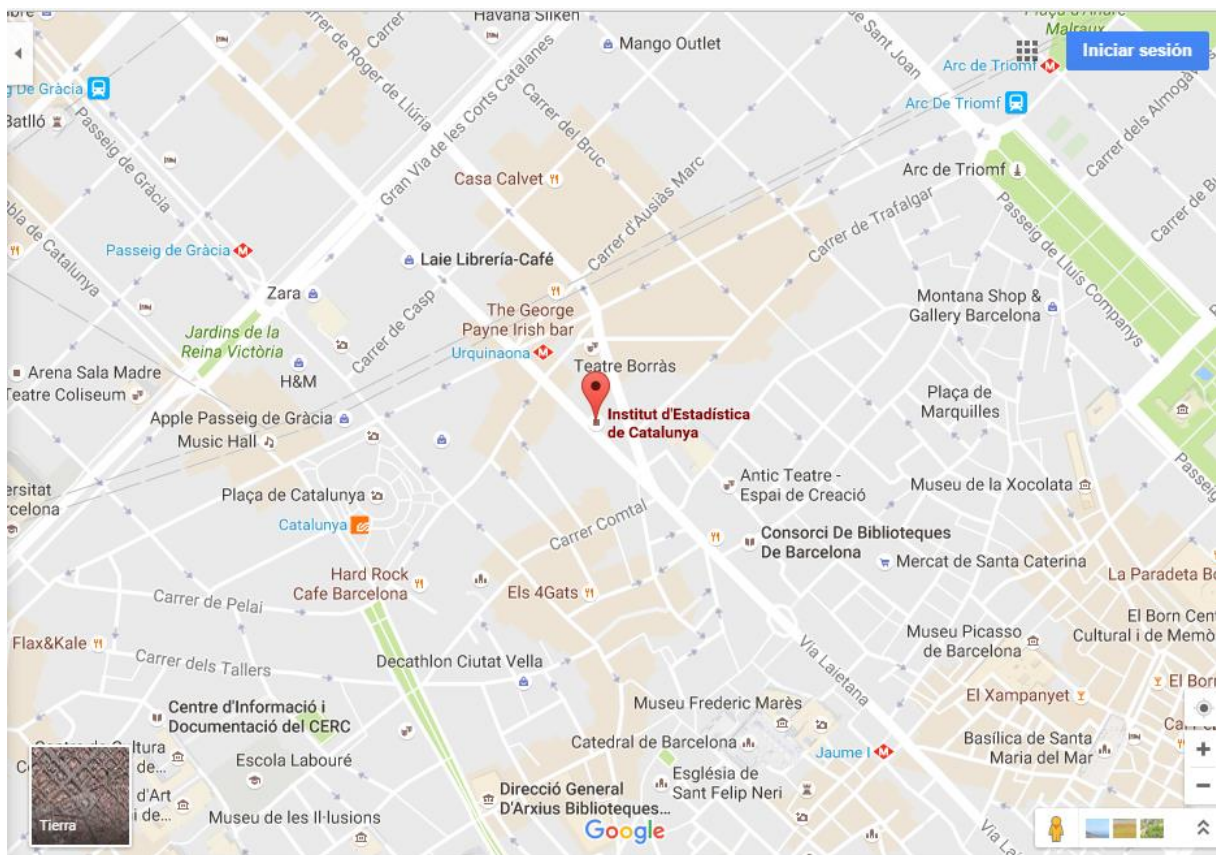
October, 17th: Deadline for registration

University of Barcelona, December 2016.

Location:

Events Room of the Statistical Institute of Catalonia – IDESCAT.

Via Laietana, 58, 08003 Barcelona, Spain.



Sponsors:



Sociedad de Estadística e Investigación Operativa

<http://www.seio.es/>



Generalitat de Catalunya
**Institut d'Estadística
de Catalunya**

<http://www.idescat.cat/>



**SOCIETAT CATALANA
D'ESTADÍSTICA**

<http://soce.iec.cat/>



Servei d'Estadística
Universitat Autònoma de Barcelona

<http://sct.uab.cat/estadistica/>

Annual meeting of the SEIO Working Group in Multivariate Analysis and Classification (AMyC), Barcelona, October 20–21, 2016:

The II International Workshop on Proximity Data, Multivariate Analysis and Classification has taken place during October 20-21, 2016 in Barcelona, Spain:

<http://www.ub.edu/wamyc/>.

It has been organized by the Multivariate Analysis and Classification Spanish SEIO Group AMyC,

<http://amyc.seio.es/>.

The Spanish Group of Multivariate Analysis and Classification is a Working Group of more than 50 researchers from all the Spanish universities. Every year, the Working Group organizes a meeting to promote the communication between its members and between them and other researchers, and to contribute to the development of the Multivariate Analysis and Classification field and related problems and applications. The I International Workshop on Proximity Data, Multivariate Analysis and Classification took place in Granada, October 2014. The web of the first edition of the workshop was <http://www.ugr.es/~amyc/EVENTS/WAMYC/>.

The founder of the Spanish Group AMyC is **Carles M. Cuadras**. This second edition of the workshop is especially dedicated to give a tribute to Dr. Cuadras for his relevant contributions to multivariate analysis and his trajectory within the working Group AMyC.

The topics of interest of the Workshop comprise any related problem to Multivariate Analysis and Classification both from a theoretical or a computational point of view, and their applications. It also includes problems related to unsupervised or supervised statistical learning related to big data analysis.

Reduced Program:

	<i>Thursday, 20</i>	<i>Friday, 21</i>
8.30 – 9 h	<i>Registration</i>	<i>Registration</i>
9 – 10 h	<p><i>Opening</i></p> <p><i>José Fernando Vera, Eva Boj and Frederic Udina</i></p> <p><i>Invited Inaugural Talk</i></p> <p><i>Carles M^a Cuadras</i></p>	<p><i>Plenary Invited Talks</i></p> <p><i>Chair: Eva Boj</i></p> <p><i>- José Fernando Vera</i></p> <p><i>- Daniel Oberski</i></p> <p><i>- José Luis Vicente-Villardón</i></p> <p><i>- Eva Boj</i></p>
10 – 11 h	<p><i>Special Session:</i></p> <p><i>Former Students of Dr. Cuadras 1</i></p> <p><i>Chair: Josep Fortiana</i></p>	
11 – 11.15 h	<i>Coffee Break</i>	<i>Coffee Break</i>
11.15 – 12.15 h	<p><i>Special Session:</i></p> <p><i>Former Students of Dr. Cuadras 2</i></p> <p><i>Chair: Josep Fortiana</i></p>	<p><i>Poster Session / Regular Session</i></p> <p><i>Chair: José Luis Vicente-Villardón</i></p>
12.15 – 13 h	<p><i>Invited Talk</i></p> <p><i>Chair: José Fernando Vera</i></p> <p><i>Invited Speaker: Elías Moreno</i></p>	<p><i>Invited Talk: IDESCAT</i></p> <p><i>Chair: Cristina Rovira</i></p> <p><i>Invited Speaker: Mònica Gasulla</i></p>
13 – 14 h	<p><i>Special Session: Compositional Data</i></p> <p><i>Chair: Josep Antoni Martín</i></p>	<p><i>Special Session: GRBIO</i></p> <p><i>Chair: Lupe Gómez</i></p>
14 – 16.30 h	<i>Lunch</i>	<i>Lunch</i>
16.30 – 19 h	<p><i>Special Session: Statistics for Automated Mass-Crowd Counting</i></p> <p><i>Organizer: Albert Satorra</i></p>	<p><i>AMyC Group Meeting</i></p> <p><i>Closing</i></p>
16.30 – 17.30 h	<p><i>Invited Talk</i></p> <p><i>"A computer vision method for image crowd counting"</i></p> <p><i>Haroon Idrees</i></p> <p><i>Department of Computer Vision, University of Central Florida</i></p>	

17.30 – 17.45 h	<i>Coffee Break</i>	
17.45 – 19 h	<p><i>Contributed Talks</i></p> <p><i>Chair: Albert Satorra</i></p> <ul style="list-style-type: none"> - Marcos Cruz, Javier González-Villa, Luis Manuel Cruz-Orive: "Efficient and unbiased crowd size estimation" - Llorens Badiella: "Attendance estimation for massive demonstrations using basic statistics" - Albert Satorra, Josep M. Oller: "Crowd size estimate based on line-density integration and computerized count of images" 	
19.15 – 19.45 h	<p><i>Invited Talk</i></p> <p><i>Chair: Carles M^a Cuadras</i></p> <p><i>Invited Speaker: Daniel Peña</i></p>	

Program:

Thursday, 20

8.30 – 9 h Registration.

9 – 10 h *Opening.*

José Fernando Vera (University of Granada), Eva Boj (University of Barcelona) and Frederic Udina (Statistical Institute of Catalonia, IDESCAT).



Frederic Udina (on the left) and Carles M. Cuadras (on the right).

Invited Inaugural Talk

COMPARING MULTIPLE FACTOR ANALYSIS WITH RELATED METRIC SCALING FOR REPRESENTING MULTIVARIATE DATA.

Invited Speaker: Carles M. Cuadras, University of Barcelona, Spain

ABSTRACT

After considering two different distributions with similar moment generating functions, which can be distinguished by using characteristic functions, we perform a comparison of four methods of multidimensional scaling on the basis of mixed variables, with two or more related data sets. The methods are: metric scaling with Gower's similarity, multiple factor analysis (MFA), related metric scaling (RMDS) and canonical analysis in MANOVA. We specially compare MFA and RMDS. Although both methods provide similar results, because the first principal component expresses dependence, RMDS is less redundant and performs better if the data sets have common principal dimensions.

10 – 11 h *Special Session: Former Students of Dr. Cuadras 1*

Chair: Josep Fortiana, University of Barcelona, Spain.

11 – 11.15 h *Coffee Break*

11.15 – 12.15 h *Special Session: Former Students of Dr. Cuadras 2*

Chair: Josep Fortiana, University of Barcelona, Spain.

Participants in the Special Sessions of Former Students of Dr. Cuadras:

Lupe Gómez, Jordi Ocaña, Daniel Cuadras, Conchita Arenas, Josep M^a Oller, Martín Ríos, Josep M^a Duran, Joan Manel Batista, Purificación Galindo and Josep Fortiana.

See the **Appendix** for the contributions of the participants in the Special Sessions of Former Students of Dr. Cuadras.

12.15 – 13 h *Invited Talk*

Chair: José Fernando Vera, University of Granada, Spain.

PROBABILISTIC CLUSTERING: ASYMPTOTIC RESULTS.

Invited Speaker: Elías Moreno, University of Granada, Spain

ABSTRACT

Clustering k samples is a statistical model selection problem, and modeling approaches include mixture models and product partition models. Here we develop an asymptotic analysis for the product partition models approach.

A Bayesian viewpoint is adopted and posterior model consistency of the clustering procedure is studied. We examine the asymptotic behavior of the posterior model distribution mainly for four priors on the model space, the hierarchical uniform prior, the Ewens-Pitman prior, the Janssen-Lui prior and the uniform prior.

We have found that while the choice of the prior on the model parameters is not relevant for the consistency of the procedure, the choice of the prior on the model space is of utmost importance, almost overshadowing the other parts of the clustering problem. The analysis indicates that the Janssen-Lui prior and the uniform prior are not appropriate priors for clustering.

13 – 14 h *Special Session: Compositional Data*

Chair: Josep Antoni Martín, University of Girona, Spain.

SPURIOUS CORRELATION IN COMPOSITIONAL DATA ANALYSIS: FROM PEARSON TO COPULAS, THROUGH SPEARMAN AND KENDALL

Speaker: Maribel Ortego, Universitat Politècnica de Catalunya – BarcelonaTECH, Spain.

Co-author: Juan José Egozcue

ABSTRACT

A common issue in statistical applications is the modelisation of dependence between two or more variables. Practitioners often use the Pearson correlation coefficient to measure dependence,

although this coefficient only captures linear dependence. Other coefficients such as Kendall's τ or Spearman's ρ , among others Scarsini (1984); Schweizer and Wolff (1981) are popular alternatives for those wanting to model dependencies other than linear. These coefficients are linked with copula functions Genest and Favre (2007); Nelsen (1999). As copulas seem a useful tool to represent dependence between variables, these coefficients are an attractive alternative.

The representation of dependence is a key issue when dealing with compositional data Pawlowsky-Glahn et al. (2015). In fact, the first efforts towards the suitable treatment of compositional data were driven by the burst of spurious (Pearson) correlation into the scene. Although Pearson correlation is spurious for compositional data, a simple log-ratio treatment is enough for the avoidance of this problem. Some researchers may expect that Kendall's τ , Spearman's ρ or the copula between parts of a composition are not spurious thus suggesting a way to circumvent log-ratio treatment. This simplification should be avoided.

In this work, Pearson, Kendall and Spearman's dependence coefficients and copula functions are used to measure the dependence between parts of a composition and its subcompositions. It is shown that if the compositional structure of data is ignored, the use of these measures of dependence also presents great shortcomings, as they are subcompositionally incoherent and, consequently, they are spurious measures of dependence between parts of a composition.

References

- Genest, C. and Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4):347–368.
- Nelsen, R. B. (1999). *An introduction to copulas*. Springer-Verlag, New York, NY, USA. 216p.
- Pawlowsky-Glahn, V., Egozcue, J. J., and Tolosana-Delgado, R. (2015). *Modeling and analysis of compositional data*. Statistics in practice. John Wiley & Sons, Chichester UK. 272 pp.
- Scarsini, M. (1984). On measures of concordance. *Stochastica*, 8(3):201–218.
- Schweizer, B. and Wolff, E. F. (1981). On nonparametric measures of dependence for random variables. *The Annals of Statistics*, 9(4):870–885.

A PARAMETRIC APPROACH TO COUNT DATA IN COMPOSITIONAL DATA SETS

Speaker: Marc Comas-Cufí, University of Girona, Spain.

Co-authors: Josep A. Martín-Fernández, Glòria Mateu-Figueras, Javier Palarea-Albaladejo.

ABSTRACT

Compositional data (coda) are multivariate positive observations representing relative contributions to a total. The fact that the data are understood as relative amounts is commonly reflected by expressing them in percentage or similar units. Typical examples include chemical or nutritional compositions, time use budgets and so on. The log-ratio methodology has become the mainstream approach to tackle their statistical analysis. Basically, it assumes the relevant information is contained in the ratios between the parts. However, the use of ratios meets practical difficulties when the observed coda contain zeros. Hence, data pre-processing frequently involves the imputation of zeros by sensible values. A number of proposals have been introduced for the case in which zeros occur in continuous coda as a consequence of a censoring problem.

The equivalent zero problem in the context of discrete compositional count data has been only recently approached from a non-parametric point of view. Compositional count data are discrete vectors accounting for the number of outcomes falling into an array of mutually exclusive categories. A compositional analysis is deemed appropriate when the total sum of those vectors is not of interest.

Zero values usually happen in the most infrequent categories as a result of insufficient number of trials. We propose a parametric count zero replacement based on compounding a log-ratio normal distribution defined on the simplex, the sample space of compositional data, and a multinomial distribution. The log-ratio normal is used to model the vector of probabilities of the categories, whereas the multinomial deals with the counts. Parameter estimation is conducted by maximum likelihood through a Monte Carlo implementation of the Expectation-Maximisation (EM) algorithm.

The estimated multinomial probabilities can be used to impute the count zeros without altering the relative relationships structure of the components. The performance of the method is illustrated by real and simulated data.

LIQUIDITY AND DEBT PROFILES. FROM CLASSIFICATION ACCORDING TO FINANCIAL RATIOS TO COMPOSITIONAL CLASSIFICATION.

Speaker: Germà Coenders, University of Girona, Spain.

Co-author: Salvador Linares Mustarós

ABSTRACT

Financial ratios (Chen and Shimerda, 1981; Walsh, 2008) have as main objective to eliminate the influence company size in the study of its financial and economic structure. From the classic works of Altman (1968) and Beaver (1966) there is a consensus on the fact that the relative financial statement magnitudes are those that best predict many variables of interest. Classic examples are bankruptcy (Altman, 1968; Ohlson, 1980) and the survival time of companies (Chen and Lee, 1993). Ratios are also used to classify companies from the similarity of their financial structures (Cinca, Molinero and Larraz, 2005; Cowen and Hoffer, 1982, Yoshino and Taghizadeh-Hesary, 2015). The latter use is the subject of this paper.

We present a simplified balance sheet structure that allows us to illustrate some of the most common ratios. We focus on liquidity and debt ratios indicating if the firm can deal with longer or shorter term debts, with what is immediately available as cash, or can be turned into cash within a shorter or longer term (accounts receivable and inventory). The simplified asset contains, from lower to higher liquidity, x_1 = Fixed assets; x_2 = Inventory; x_3 = Accounts receivable; and x_4 = Cash and equivalents. The simplified liabilities/equity contain x_5 = Equity, x_6 = Long-term liabilities; x_7 = Short-term liabilities.

Many ratios compare one or more assets with one or more equity/liability parts. For example, the cash ratio is:

$$x_4 / x_7$$

Ratios have statistical problems related to their lack of symmetry, which prevents the use of models based on normal distributions (Barnes, 1982). They often have non-linear relationships and outliers (Cowen and Hoffer, 1982). Ratios are not symmetrical with respect to unit value indicating equality between both quantities. In our cash ratio, for example, the distance from the unit is not the same when x_7 doubles x_4 and when x_4 doubles x_7 . This may compromise the interpretation of distances among companies calculated from financial ratios, and therefore classification. Reformulating the cash ratio as a log-ratio solves these problems at once.

$$\ln (x_4 / x_7)$$

Analysis of log-ratios is the standard methodology in the analysis of compositional data (Aitchison,

1986; Pawlowsky-Glahn, Egozcue and Tolosana-Delgado, 2015). Compositional data are those that contain only information about the relative magnitude of the parts of a whole (eg in%), and usually have a restricted sum (for example, $S = 100$). In the case of balance sheets, two restrictions, because of the equality between assets and liabilities.

$$x_1 + x_2 + x_3 + x_4 = x_5 + x_6 + x_7 = S$$

Thus, 5 nonredundant log-ratios contain all the information about the ability to pay debt, and this is the number used in the analysis of compositional data. The financial literature warns of the mutual redundancy of more than 100 commonly used ratios, especially those that have a common denominator, which leads to spurious correlations (e.g., Lev and Sunder, 1979). Therefore, researchers usually seek to reduce their numbers for the sake of parsimony (Pindado and Rodrigues, 2004), for example by using factor analysis techniques (eg, Boyacioglu, Kara and Baykan, 2009; Cowen and Hoffer, 1982; Yoshino and Taghizadeh-Hesary, 2015). Compositional data analysis makes this step unnecessary.

In this paper we propose to treat assets and liabilities/equity as two separate compositions, and separately apply the so-called isometric log-ratio transformation (Egozcue, Pawlowsky-Glahn, Mateu-Figueras and Barcelo-Vidal, 2003). This achieves:

- 1) A representation of the relative magnitudes of the liability/equity parts. The isometric log-ratios compare each part in an ordered composition with the geometric mean of the previous parts. The first isometric log-ratio compares x_5 with x_6 , and the second x_7 with the geometric mean of x_5 and x_6 .
- 2) A representation of the relative asset magnitudes, with three isometric log-ratios calculated in the same way.
- 3) Both representations also contain information on the relative magnitude of any asset compared to any liability/equity. For example, subtracting the last log-ratios of assets and liabilities/equity leads to comparing x_4 and x_7 , like the cash ratio does.
- 4) The Euclidean distance between two companies calculated on the log-ratios above equals the sum of the compositional Aitchison distances between the two companies calculated separately from assets and liabilities/equity.
- 5) Once data have been transformed, a fully standard cluster analysis algorithm can be applied to

classify companies according to their financial structure. For example, Ward's method on the distances, or the k-means method on the log-ratios.

An illustration of the procedure is provided displaying distinct profiles of liquidity and debt problems.

14 – 16.30 h *Lunch*

Restaurant "Els Quatre Gats". C/ Montsió 3. 08002_Barcelona.

16.30 – 19 h *Special Session: Statistics for Automated Mass-Crowd Counting*

Organizer: Albert Satorra, Pompeu Fabra University, Spain.

DETAILS of the Special Session: Statistics for Automated Mass-Crowd Counting

16.30 – 17.30 h *Invited Talk*

Chair: Albert Satorra, Pompeu Fabra University, Spain.

A COMPUTER VISION METHOD FOR IMAGE CROWD COUNTING.

Invited Speaker: Haroon Idrees, Department of Computer Vision, University of Central Florida, USA.

ABSTRACT

In this talk, I will present our recent work on visual analysis and counting of extremely dense crowds. Since computer vision algorithms require labeled training data, I will begin by describing the process of collecting crowd images followed by their manual annotation using a GUI-based tool. The first algorithm we developed in 2013 relies on multiple sources of information such as low confidence head detections, repetition of texture elements (using SIFT), and frequency-domain analysis to estimate counts, which is followed by a global consistency constraint on counts using Markov Random Field. Next, I will present our approach using 'deep Learning' which attempts to mimic animal brain through Artificial Neural Networks. The experiments performed on a recent extension of a dataset with one million humans shows the power of this approach over existing machine learning algorithms. For detecting and localizing humans in crowd images, I will discuss

how we explore context in the form of locally-consistent scale prior which captures the similarity in scale in local neighborhoods. Using the scale and confidence of detections obtained from an underlying human detector, we infer the scale prior in an iterative mechanism. The detections obtained are then reasoned for occlusion using Binary Integer. Compared to counting approaches which only give a single estimate for entire image, human detection gives locations of humans as well. Finally, I will conclude the talk with the counting results on images from Catalan Separatist Movement demonstrations which took place on Sept 11, 2015 and Sept 11, 2016.

BIOGRAPHY:



Haroon Idrees received the BS degree in Computer Engineering from Lahore University of Management Sciences, Pakistan, and received his PhD in Computer Science in 2014 from the University of Central Florida, United States. He is currently working as a Postdoctoral Associate with Center for Research in Computer Vision at the University of Central Florida. His research interests include crowd analysis, object detection, action recognition, visual tracking, multi-camera and airborne surveillance, deep learning and multimedia content analysis. He has published several papers in top computer vision journals and conferences such as IEEE Transactions on Pattern Analysis and Machine Intelligence, Computer Vision and Pattern Recognition (CVPR), European Conference on Computer Vision (ECCV) and Image and Vision Computing. He has also served as the program chair for THUMOS Action Recognition Challenge Workshop, CVPR 2015, as well as Multiple Object Tracking (MOT) Challenge Workshop, ECCV 2016.

17.30 – 17.45 h *Coffee Break*

17.45 – 19 h *Contributed Talks*

Chair: Albert Satorra, Pompeu Fabra University, Spain.

EFFICIENT AND UNBIASED CROWD SIZE ESTIMATION

Speaker: Marcos Cruz, University of Cantabria, Spain.

Co-authors: Javier González-Villa, Luis Manuel Cruz-Orive.

ABSTRACT

Crowd size estimation is a long standing problem. Exhaustive manual counting is tedious, slow, difficult to verify and unfeasible for large populations. Automatic detection and counting algorithms are generally biased and usually have high and unpredictable variance. An alternative is to multiply population density with some reference area but, unfortunately, sampling details, handling of edge effects, etc., are seldom described, whereby the statistical properties of the estimators are generally unknown. We address the problem by using principles of geometric sampling. These principles are old and solid, but largely unknown outside the areas of microscopy and stereology. The proposed design is design unbiased irrespective of population size, pattern, perspective artifacts, etc. The implementation is very simple—it is based on the random superimposition of coarse quadrat grids. The essential ingredient is proper, well defined sampling. We propose (and check via Monte Carlo resampling) a new theoretical variance prediction formula. As far as efficiency, for the different cases studied counting about 50 (100) individuals in 20 quadrats, yielded relative standard errors of about 8% (5%) with a few minutes work. This fact may effectively break the barrier hitherto imposed by the current lack of reliable automatic detection algorithms, because semiautomatic sampling and manual counting may be an attractive option.

ATTENDANCE ESTIMATION FOR MASSIVE DEMONSTRATIONS USING BASIC STATISTICS

Speaker: Llorens Badiella, Autonomous University of Barcelona, Spain.

ABSTRACT

Whenever there is a demonstration, the same dilemma arises: what was the number of participants? In order to calibrate the relevance of the concentration it is mandatory to use reproducible and valid methods.

Aerial image processing techniques are very promising in order to accomplish the counting task. These methods, especially in very large demonstrations, require to be calibrated by field work. Moreover, in most of the occasions, there is no availability of such kind of images. For these reasons, field measurements are still necessary.

Jacobs (1967) proposed a simple method: measuring the total area occupied by the demonstration and multiplying it by a density estimate. This method can be used in large demonstrations by splitting the total area into nearly dense portions and estimating each parcel. This method has several limitations such as dynamic or over crowded concentrations. The most critical issue is that it does not provide estimation variability.

The author proposes a method based on obtaining a number of measurements performed along the concentration (or at random), counting all participants in a slice of a given width (for example 1m). This method provides valid and reproducible estimates along with its variability. However it is only applicable in static, stretched and moderately dense demonstrations.

CROWD SIZE ESTIMATE BASED ON LINE-DENSITY INTEGRATION AND COMPUTERIZED COUNT OF IMAGES

Speaker: Albert Satorra, Pompeu Fabra University, Spain.

Co-author: Josep M. Oller

ABSTRACT

Protest movements, sometimes promote concentrations of people, in technical terms, the so-called crowds. Since the size of the crowd (SC) tends to be used as the measure of the strength of the protest movement, heated debates arise around the truth of the SC reported by the different agencies (organizers, government, newspapers, etc.). In words of Watson and Yip (2011) "There is a large amount of variability in crowd estimates [SC] for two reasons: it is difficult to do, and there are strong motivations for getting it wrong!" (p. 104). For example, in the Separatist Movement demonstrations of 11 Sept 2015 in Barcelona, the organizers claimed 2 million people while other agencies provided a SC estimates around 1/2 million people. Such huge disparities among SC estimates evidence the presence of "political" (non-statistical!) bias in some, or maybe all, of the SC estimates.

In a static crowd, the basics for computing SC is the area method formula $SC = A \times D$, where A is the area occupied by the crowd and D is a mean density estimate (people per square meter). In the case of a line-stretched crowd (LC), when the concentration takes place along an avenue of a city, then the area formula is $A = L \times W$, where L is the length of the concentration, and W is the mean width along the concentration. In an LC, the parameter of length L tends to be widely agreed, but the parameters W of the width and D of density typically remains in dispute.

In this talk, we describe a SC estimate for a static LC. We present a method that makes use of the usually undisputed length parameter L but has the advantage of circumventing the controversial parameters of width W and density D. The key data for the new method are computerized counts of aerial images taken along the main axis of the crowd. The SC estimate is then produced by integrating the line density estimated from the automatic counts of images. Since integration is mathematical, and the counting of the images is computerized, non-statistical (i.e., political) bias is avoided by this CS estimate. The method is illustrated in the crowd analysis of the Catalan Separatist Movement demonstrations which took place on Sept 11, 2015 and Sept 11, 2016.

19.15 – 19.45 h *Invited Talk*

Chair: Carles M^a Cuadras, University of Barcelona, Spain.

AN UNIFIED PROCEDURE FOR CLUSTERING TIME SERIES

Invited Speaker: Daniel Peña, Carlos III University of Madrid, Spain.

Co-author: Andrés M. Alonso

ABSTRACT

We present a new way to find clusters in large vectors of time series. The procedure has two steps. In the first series are split by their dependency, which is measured by the determinant of their correlation matrix including lags. Second, inside each groups the series are split by putting together series with a similar autocorrelation structure. The procedure is automatic and can be applied to large data sets. Several examples are presented to illustrate the procedure.

Friday, 21

8.30 – 9 h Registration.

9 – 11 h *Plenary Invited Talks*

Chair: Eva Boj, University of Barcelona, Spain.

ON THE ADDED VALUE OF CLUSTERING IN MULTIDIMENSIONAL SCALING

Invited Speaker: José Fernando Vera, University of Granada, Spain.

ABSTRACT

The principal aim in Multidimensional Scaling is the representation of a set of objects usually in a Euclidean space of low dimension, when the information is given in terms of dissimilarities. To enhance the interpretation of the MDS solution and/or to obtain an adequate fit of the model when the number of objects is large for their representation, cluster and MDS methods have been used in conjunction to represent the information of similar proximity data by means of groups. The data are categorized into a small number of groups of similar elements, such that each group label summarizes the required information about the group.

Although some clustering methods also provide a spatial representation of clusters, this will be low dimensional only, as e.g. in K-means, when the number of clusters and the original dimension of datasets are not larger than two or three, and the information is not given in terms of dissimilarities. Thus, a usual practice is to reduce the number of dimension by MDS, and then superimposing the results on a clustering analysis on this representation. Nevertheless, it should be noted that the reduced space is optimal for the embedded points but not for the superimposed clusters, while the cluster structure is optimal in the original, non-reduced space.

An advisable methodology is the combination of classification and spatial representation in a procedure that simultaneously makes clustering, while not the original objects but the cluster centres are represented in a low dimensional space. This method has the advantage that the objects or individual information is summarized by means of groups, which significantly reduces the number of parameters to be estimated in the model. Several of these combined procedures have

been proposed in a deterministic framework for two-way data (Vera, Macías and Angulo 2008), but also for preference data (Vera, Macías and Heiser, 2013). In a probabilistic framework, the combination of latent class models and Simulated Annealing have been proposed for MDS (Vera, Macías and Heiser, 2009a; Vera Macías and Angulo, 2009) and Unfolding (Vera, Macías and Heiser, 2009b), which also provides statistical tools to determine both the number of clusters and the dimensionality of the model.

Vera, J.F., Macías, R. and Angulo, J.M. (2008). Non-stationary spatial covariance structure estimation in oversampled domains by cluster differences scaling with spatial constraints. *Stochastic Environmental Research and Risk Assessment* 22, 95–106.

Vera, J.F., Macías, R. and Angulo, J.M. (2009). A latent class MDS model with spatial constraints for non-stationary spatial covariance estimation. *Stochastic Environmental Research and Risk Assessment*. 23(6), 769–779.

Vera, J.F., Macías, R. and Heiser, W.J. (2009a). A latent class multidimensional scaling model for two-way One-mode continuous rating dissimilarity data. *Psychometrika* 74(2), 297–315.

Vera, J.F., Macías, R. and Heiser, W.J. (2009b). A dual latent class unfolding model for two-way two-mode preference rating data. *Computational Statistics and Data Analysis* 53(8), 3231–3244.

Vera, J.F., Macías, R. and HEISER, W.J. (2013). Cluster differences unfolding for two-way two-mode preference rating data. *Journal of Classification* 30(3), 370–396.

BAYESIAN MIXTURE PRIORS FOR MEASUREMENT INVARIANCE STRUCTURAL EQUATION MODELING

Invited Speaker: Daniel Oberski, Utrecht University, the Netherlands.

ABSTRACT

Multiple group latent variable models often have a large number of parameters that may be similar across groups or radically different. Examples include cross-country analyses of mathematics ability such as PISA, or sociological analyses comparing questionnaire results from the European Social Survey. The goal of such analyses is usually to compare groups, which requires the groups share some measurement parameters. This puts the researcher into a difficult position: allowing for all possible cross-group differences in measurement parameters leads to underidentification, while

restricting all parameters to be shared across the groups may lead to bias in the analyses. Moreover, a specification search using techniques such as the score tests (modification index) may be prohibitively complex in practice.

A solution to these issues suggested by Muthen & Asparouhov (2012) is Bayesian analysis including strong priors on the cross-group differences in measurement parameters. However, simulation studies (e.g., Van de Schoot et al 2014) have shown that this approach does not work in practice, because it is unable to deal with large deviations in measurement parameters without disabling the advantages of parameter sharing. Moreover, suggested techniques to detect such deviations using posterior predictive checks have not solved this problem.

In this presentation I discuss an alternative solution, which can be seen as a generalization of the priors suggested by Muthen & Asparouhov. In particular, I discuss a family of mixture priors for cross-group differences in measurement parameters that have been designed to allow for small differences over groups, which also accounting for large differences when they exist. A common framework including several existing "measurement invariance" or "differential item functioning" models is discussed, as well as some preliminary simulation results. It is hoped future work will facilitate practical research that aims to compare latent variable models across a large number of groups.

LOGISTIC BIPLOTS FOR BINARY, NOMINAL AND ORDINAL DATA

Invited Speaker: José Luis Vicente-Villardón, University of Salamanca, Spain.

Co-author: Julio César Hernández-Sánchez

ABSTRACT

Biplots are simultaneous representations of rows and columns of a data matrix. Biplots (or related methods) are becoming one of the most popular techniques for visualizing data sets, especially when dealing with continuous data. Recently we have proposed biplot representations for categorical (binary, nominal and ordinal) based on logistic response models. The coordinates of individuals and variables are computed to have logistic responses along the biplot dimensions. The methods are related to logistic regression in the same way that Classical Biplots are related to linear regression, thus we refer to the methods as (Binary, Nominal or Ordinal) Logistic Biplots. In the same way as Linear Biplots are related to Principal Components Analysis, Logistic Biplots are related

to Latent Trait Analysis or Item Response Models. The methods are illustrated with two data sets: one taken from the literature and another one with data from a study of job satisfaction of doctorate (PhD) holders in Spain.

Keywords: Logistic Biplot, Binary, Nominal and Ordinal Data.

References

Vicente-Villardón, J. L., Galindo, M. P. and Blazquez, A. (2006). Logistic Biplots. In: *“Múltiple Correspondence Análisis And Related Methods”*. Grenacre, M & Blasius, J. Eds. Chapman and Hall. Boca Ratón.

Demey, J., Vicente-Villardón, J. L., Galindo, M.P. and Zambrano, A. (2008) Identifying Molecular Markers Associated With Classification Of Genotypes Using External Logistic Biplots. *Bioinformatics*, 24(24), 2832-2838.

Hernández-Sánchez, J. C., and Vicente-Villardón, J. L. (2016). Logistic biplot for nominal data. *Advances in Data Analysis and Classification*, 1-20. doi:10.1007/s11634-016-0249-7.

Vicente-Villardón, J. L., and Hernández-Sánchez, J. C. (2014). *Logistic Biplots for Ordinal Data with an Application to Job Satisfaction of Doctorate Degree Holders in Spain*. arXiv preprint arXiv:1405.0294.

DISTANCE-BASED GENERALIZED LINEAR MODELS: FROM OLS TO IWLS

Invited Speaker: Eva Boj, University of Barcelona, Spain.

ABSTRACT

In this invited talk the performance of a recently developed R package called *dbstats* and available at <http://CRAN.R-project.org/package=dbstats> from the Comprehensive R Archive Network is showed (see [1]). *dbstats* contains classes and functions implementing distance-based prediction methods such as distance-based linear model (DB-LM), local DB-LM, distance-based generalized linear model (DB-GLM), local DB-GLM and distance-based partial least squares regression (DB-PLSR) (see [2] and [3]). First, the main features of DB-LM and of DB-GLM are reviewed. DB-GLM is an extension of DB-LM to the framework of GLM, where the iterative least squares, IWLS, procedure is used instead of the ordinary least square, OLS, estimation. Second, the *dbstats* package and its functions *dblm* and *dbglm* are presented. Finally, examples to illustrate the distance-based

regression models and the usage of the package are made. DB-LM and DB-GLM have been proposed in actuarial applications as are: pricing, credit scoring and claim reserving.

Keywords: dbstats, distance-based prediction, generalized linear model, actuarial sciences.

References:

[1] Boj, E., Caballé, A., Delicado, P., Fortiana, J. (2014). *dbstats: distance-based statistics (dbstats)*. R package, version 1.4.

[2] Boj, E., Delicado, P., Fortiana, J. (2010). Local linear functional regression based on weighted distance-based regression. *Computational Statistics and Data Analysis* 54, 429–437.

[3] Boj, E., Delicado, P., Fortiana, J., Esteve, A., Caballé, A. (2016). Global and local distance-based generalized linear models. *TEST* 25, 170–195.

11 – 11.15 h *Coffee Break*

11.15 – 12.15 h *Poster Session / Regular Session*

Chair: José Luis Vicente-Villardón, University of Salamanca, Spain.

Poster Session:

DB-DISCRIMINANT: AN EXTENSION

Authors: Itziar Irigoien, Francesc Mestres and Conchita Arenas (University of Barcelona).

ABSTRACT

Discriminant analysis is used by researchers in a wide variety of fields. In Cuadras et al. 1997, a distance-based discriminant rule (DB-discriminant) was introduced. We present an extension of this classifier using a depth function. The new discriminant rule combines the DB-rule and the maximal depth classifier, obtaining a rule that is often more accurate than both methods separately. To demonstrate its effectiveness the new classifier was compared with the DB-rule and the k-nearest neighbor classification method. The results show the good performance of the new classifier.

CHANGES IN THE OCCURRENCE OF EXTREME WAVE STORMS IN THE CATALAN COAST: A BAYESIAN APPROACH

Authors: Jesús Corral-López (Universitat Politècnica de Catalunya – BarcelonaTECH), Raquel Feixa-Compte and Maribel Ortego.

ABSTRACT

When dealing with historical data, a common issue for practitioners is the possible existence of trends in the series that should be taken into account. In particular, it may be interesting to analyse whether the intensity of extreme adverse weather conditions in the Catalan coast have increased over the last 60 years, probably as a consequence of climate change. A dataset has been selected to this end. For the period 1958-2015, a set of 15 different SIMAR points located in the Mediterranean coast have been selected. Information has been provided by Puertos del Estado (responsible for the coordination and control of the spanish port system's effectiveness). A set of five categories of storms has been defined, based on the values of $\log(H_s)$, the significant waveheight. For each year, storms have been classified in these categories. In order to estimate the probabilities of each category, a methodology that combines the Aitchison geometry for probabilities and Bayesian techniques has been developed.

The classification of storms into different categories provides compositional data. The sample space of compositional data is the Simplex, and it can be characterized using the Aitchison geometry. The transformation of this compositional data into ilr coordinates allows us to work in the real space. As the number of extreme weather conditions is low, and in some cases zero, it is advisable to use Bayesian statistical methods. Bayesian methods have some advantages in front of other traditional approaches. However, a common drawback of the use of Bayesian techniques is that the expression of the posterior density may be complex. Nevertheless, the use of modern MonteCarlo (MC) methods, such as the Hamilton Montecarlo implemented in Stan, provide tools to solve this issue. A sample of the posterior distribution of parameters has been obtained. This distribution is the basis for the study of the evolution of extreme storms in this period of time. Finally, a regionalisation of SIMAR nodes has been performed. Nodes have been grouped in clusters according to the information provided by the samples of the posterior. A hierarchic cluster method based on a Mahalanobis-type distance has been used. This method allows grouping the nodes with more correlation, taking into account the position of each node. Following the regionalization of the nodes, the possible existence of trends in the intensity of storms has been discussed.

MEAN SQUARED ERROR FOR NEW OBSERVATIONS IN LOGISTIC REGRESSION

Authors: Eva Boj and M^a Teresa Costa (University of Barcelona).

ABSTRACT

The expression of the mean squared error of prediction for new observations is obtained when using logistic regression. The mean squared error is expressed as the sum of the process variance and of the estimation variance. To obtain the estimation variance two procedures are applied: the application of the delta method to obtain an approximated formulation, and the use of bootstrap methodology to obtain the predictive distribution. When using bootstrapping, in particular bootstrap with Pearson residuals, we are able to obtain the estimation of the distribution of the predicted value. We can approximate the estimation variance with the standard deviation of the distribution. In addition, with the distribution of each new observation confidence intervals can be calculated to help us in the knowledge of the randomness of the data and of the new predicted values. To illustrate the calculus and the usefulness of the mean squared error in logistic regression for new observations, a practical application is made with credit risk real data.

Keywords: Logistic regression, mean squared error, estimation variance, delta method, bootstrapping residuals.

Regular Session:

INTRODUCING SEMANTICS INTO THE DISTANCE-BASED METHODS. AN APPLICATION TO CLUSTERING IN THE FIELD OF TOURISM

Speaker: Karina Gibert, Universitat Politècnica de Catalunya – BarcelonaTECH, Spain.

Co-author: Aida Valls

ABSTRACT

One of the active research lines in multivariate data analysis has been the treatment of heterogeneous data matrices containing different types of variables simultaneously. Indeed, extending the multivariate analysis to data other than numerical, provides a richer expressiveness to describe individuals and opens the door to introduce into the analysis qualitative information which can be relevant in many domains (like, for example, the specy of algae contaminating a certain water reservoir, or the form of a tumour related to a cáncer disease). Most of the works in

this line are related with distance-based methods, and propose compatibility measures that enables comparison between individuals simultaneously described by different types of variables. Often, numerical and categorical (qualitative) variables are considered in this measures (Gower 1971) (Gibert 1996). Classically, the qualitative part is treated through some distance dealing with equality of categories in a more or less sophisticated form (the chi2 distance takes into account the quantity of information provided by each category). However, in the new context of data science, there are a number of applications where qualitative information is a clue part of the data sources, and a clue part of the related decisional processes. It has been seen that the semantics of the terms becomes critical to properly measure distances or similarities among terms (as an example, it seems obvious that in a variable describing the type of monument, distances between church and cathedral must be smaller than distances between church and museum). To this purpose, ontologies has been used in the literature as a reference structure formalizing the semantic relationships between terms and to be taken into account for computing distances among words. In his paper, the semantic variable is introduced as a categorical variable semantically interpreted by means of a reference ontology, the superconcept based distance is presented to compare words of a semantic variable by exploiting the structure of the reference ontology, and the generalized gibert's metrics as a compatibility measure including numerical, categorical and semantic variables simultaneously. The impact of using semantic variables in clustering methods, with respect to use classical categorical variables only is illustrated by means of a real application for characterizing the visitors of the Delta del Ebre National Park.

Acknowledgements: This research has been partially financed by the project DAMASK- Data-Mining Algorithms with Semantic Knowledge Project code: TIN2009-11005 Authors thanks to the Fundacio d'Estudis turistics Costa Daurada for the case study and contributions into the interpretation of results.

(See the end of the **Appendix** of the contributions of the Special Sessions of Former Students of Dr. Cuadras for the participation of Karina Gibert)

INFLUENCE COMPONENTS

Speaker: José M^a Gutiérrez, University of Cádiz, Spain.

Co-author: Andrés Jiménez

ABSTRACT

This paper presents a local generalization of the Principal Component Analysis (PCA) called Influence Components (IC). The idea consists of defining the influence that the sample exerts over each point in space, using a linear convex combination of the variance matrices. The results obtained will be applied to the classification process in order to determine the projections that preserve the discriminant character of the variables while reducing the space dimensions.

Keywords: Influence Components, Principal Components, Induced Dispersion.

(See the end of the **Appendix** of the contributions of the Special Sessions of Former Students of Dr. Cuadras for the participation of José M^a Gutiérrez)

12.15 – 13 h *Invited Talk: IDESCAT*

Chair: Cristina Rovira, Statistical Institute of Catalonia, IDESCAT, Spain.

DYNAMIC FACTOR MODELS FOR GDP NOWCASTING: AN APPLICATION TO THE CATALAN ECONOMY

Invited Speaker: Mònica Gasulla, Statistical Institute of Catalonia, IDESCAT, Spain.

Co-authors: Jesús Muñoz, Jonathan Jorba, Sergio Plaza.

ABSTRACT

The Statistical Institute of Catalonia (Idescat) has among their strategic objectives the improvement of the main macroeconomic estimates of the Catalan economy, where GDP plays a central role. The importance of this variable for economic analysis and for the design of economic policy justifies the need of delivering GDP estimations in real time. For this reason, Idescat is working in the implementation of new methodological issues to reduce the delivery time of GDP estimates.

In this session we will explain firstly the results of the evaluation of the main nowcasting methodologies used internationally for estimating the GDP in real-time. Secondly, we will present the option followed by Idescat, the Dynamic Factor Models, which have proved to be robust, coherent and present a high predictive power. In this context, we present the results of the

evaluation of former works of Stock and Watson (2002), Giannone et al. (2008) and Cuevas and Quilis (2012). Finally, we show the main findings for its applicability for the Catalan economy.

13 – 14 h *Special Session: GRBIO*

Chair: Lupe Gómez, Universitat Politècnica de Catalunya – BarcelonaTECH, Spain.

INTEGRATIVE ANALYSIS OF OMICS DATA USING FUNCTIONAL PROFILES

Speaker: Àlex Sánchez-Pla, University of Barcelona, Spain.

Co-authors: Miquel Salicrú, Jordi Ocaña.

ABSTRACT

With the advent of the post-genomic era, by the change of century, new technologies such as microarrays became available. These high-throughput technologies allowed to interrogate simultaneously most of the genes in the genome which led to a change of paradigm in the way that biology and particularly disease was studied. The change also affected statistics (Efron states that the 21st century could be called the "microarray age") enhancing developments in high dimensional statistics or large-scale inference between others. One decade later the data explosion keeps growing. Not only data generation has become faster and cheaper but also new technologies have become available so that, besides measuring transcriptomics, it is now possible to measure all (most, at least) of the elements participating at other levels such as the proteome, the metabolome or the regulome. Altogether this has led to the need of methods for integrative analysis of these omics data that help to jointly analyze, combine and especially interpret the growing body of data available. The analysis of the results of high-throughput experiments aiming at its biological interpretation has a short but very dynamic history in bioinformatics. In a few years, dozens of methods of tools have been developed, most of them making some type of gene set or enrichment analysis aimed at detecting which biological characteristics can be associated with the results of an omics experiment. Most of these methods work by exploiting some type of annotation database that relates the omics features -genes, proteins etc- obtained as results of the experiments with their biological meaning. The best known of such databases is the Gene Ontology (GO). The GO is a project created and maintained by a public consortium whose main goal is to produce a controlled vocabulary that can be applied to all organisms even as knowledge of gene and protein roles in cells

is accumulating and changing. It is organized around three principles or basic ontologies: Molecular Functions (MF), Biological Processes (BP) and Cellular Components (CC). Each ontology (MF, BP or CC) consists of a high number of terms or categories hierarchically related from least (top) to most (bottom) specialized characteristics. A given gene product may represent one or more molecular functions, be used in one or more biological processes and appear in one or more cellular components. This allows a powerful way to characterize a list of biological features but is not free from complication due to this many-to-many association between genes and categories. The use of GO-based analysis for the interpretation of a single list of genes or proteins obtained from an omics experiment has become a standard practice and is not discussed here.

A different but related problem to that of biological significance of a gene list is comparing two gene lists from the point of view of their biological interpretation. That is, even if their components differ, do they reflect similar or distinct biological processes? Intriguingly this topic has received much less attention in the bioinformatic literature. In previous work we proposed a method to deal with this problem which is based on projecting the genes lists to be compared into an extended level of the Gene Ontology. This projection is used to build what is called a functional profile which can be modeled using a multinomial distribution and which is the basis for the analyses. In a first approach this method was used to compare one set of genes with a reference set, as a variant of Gene Enrichment Analysis. The method was later extended to allow for two sample comparisons, and was made available as a bioconductor R package ([goProfiles](#)). Nowadays it has become normal for many experimenters to generate not one but many omics datasets for their investigations. This, again, poses the problem of their joint analysis and interpretation. While classical and modern multivariate methods have been used for the analysis and visualization of these multi-omics dataset a difficulty still appears in adding biological knowledge to the results of these analysis. In this new, richer, context an extension of difference tests to equivalence tests based on functional profiles has proven to be useful in helping to decide up to what point the biological meaning extracted from distinct omics can be considered to be equivalent or, instead, it represents distinct aspects of the biological problem in study. In this study we will present an overview on how to apply tests based on functional profiles to the integrative analysis of multiple omics data. We will illustrate this with several examples appearing in an ongoing study of transcriptomic and proteomic analysis and discuss some limitations and possible extensions of the current methods used.

Keywords: Gene Ontology, Omics Data Analysis, Data Integration.

ENHANCED PLOTTING IN DATA ANALYSIS

Speaker: Josep M. Oller, University of Barcelona, Spain.

Co-author: Esteban Vegas

ABSTRACT

We may consider different situations in data analysis which lead us to consider an underlying geometry present in our data analysis framework which allow us to compare, classify or analyze our objects of study. In all cases the distance used must measure the differences between the studied objects. For example, we can be interested to study different statistical populations identified each one as a parametric statistical model and we try to compare them with the information metric. In this well-known situation we may obtain plots representing not only our populations as a points on a manifold but also represent other functions on this manifold, the effect of individual samples in the population estimation or confidence regions to estimate each population. But, on the other hand, we may not have, or we are not directly interested in a parametric statistical model, and our starting point could be a sample space with a kernel which allow us to compare the data. In this report we are going to focus our attention in this last case, although we are aware that both cases are really deeply related. Also some examples considered can be considered from both points of view. All this methods are also useful when we need to work with data of different nature, as we shall point out in an example. Descriptive statistics is often a recommendable first step to any complex problem and, at this point, the underlying geometry plays a major role.

JOINT MODELS FOR THE ANALYSIS OF TIME-TO-EVENT DATA WITH LONGITUDINAL INFORMATION

Speaker: Carles Serrat, Universitat Politècnica de Catalunya – BarcelonaTECH, Spain.

ABSTRACT

The aim of this presentation is to review joint modelling techniques for the simultaneous analysis of time-to-event data and longitudinal time-varying data. This is an increasing area of interest in both the methodological and the applied point of view and it allows the analysis and understanding of complex systems. Among others, three main advantages of this approach are: a) it corrects the bias derived from a traditional separate analysis, b) the modelization allows to incorporate and model the between and within correlation among observations and, c) true longitudinal profiles for

endogenous covariates can be included in the relative hazard survival sub-model. The relevant benefit of these models is being able to estimate the effect of each subject-specific longitudinal profile in the hazard function for the event of interest, in an adaptive manner. In particular, subject-specific dynamic predictions, like conditional survival functions given the available longitudinal information, can be derived. Estimation procedure and implementation in R (JM and JMbayes packages) will be introduced and some illustrations will be given. Extensions to the case of multiple time-to-event variables with multiple longitudinal covariates will be also considered.

Keywords: joint modeling, shared random effects models, relative risks models.

Acknowledgements of the Special Session: GRBIO

All the speakers wish to acknowledge the Grup de Recerca en Bioestadística i Bioinformàtica (GRBIO, Grup Consolidat 2014 SGR 464 Generalitat de Catalunya), for funding part of their work.

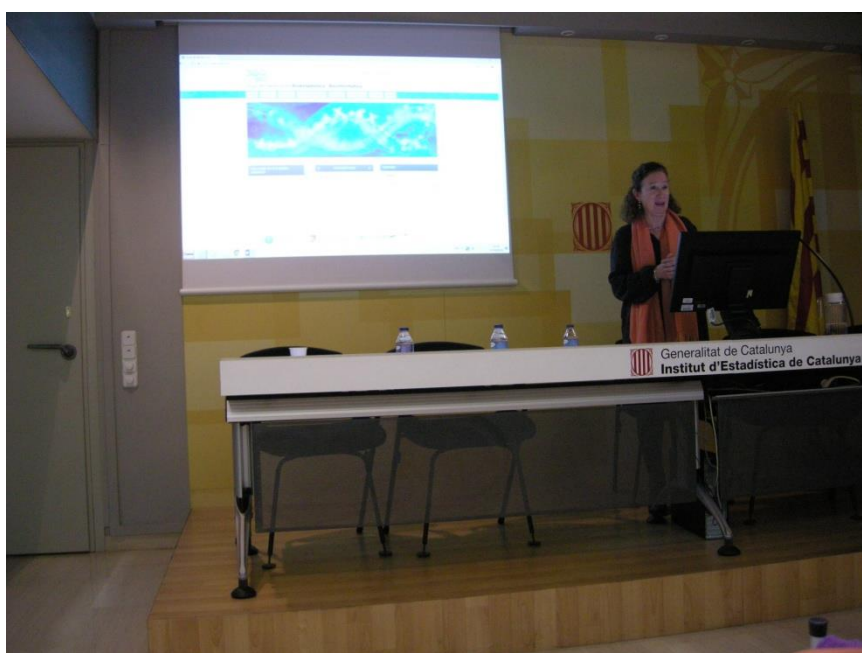
Alex Sánchez wishes to acknowledge Dr. Joan Montaner and Dr. Teresa García Berrocoso, from the Neurovascular Diseases lab at VHIR, for providing both proteomics and microarray data and Mr. Ferran Brioso of the Statistics and Bioinformatics Unit at VHIR for collaborating in the analysis of data.

14 – 16.30 h *Lunch*

16.30 – 17.30 h *AMyC Group Meeting and Closing*

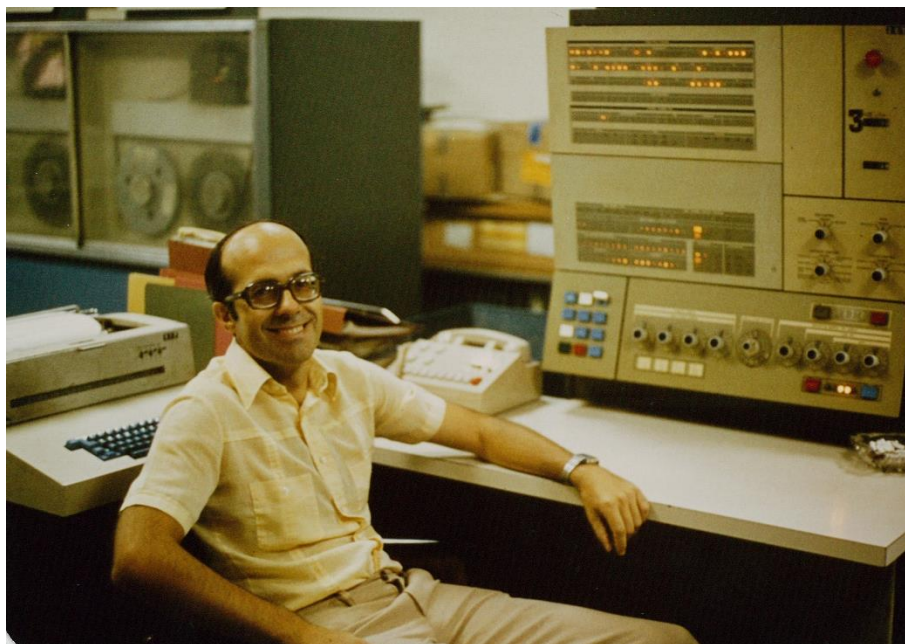
Appendix. Tribute to Dr. Cuadras

Lupe Gómez and Jordi Ocaña, Grup de Recerca en Bioestadística i Bioinformàtica (GRBIO, Grup Consolidat 2014 SGR 464 Generalitat de Catalunya), Spain.



Jordi Ocaña:

Vaig conèixer en Carles M. Cuadras a mitjans dels anys setanta del segle XX, quan ell era analista d'aplicacions al Laboratori de càlcul de la Universitat de Barcelona i feia poc que s'havia fet càrrec de la docència de l'assignatura "Biometria i Estadística" de la llicenciatura en Ciències biològiques. Un grup d'entusiastes, en la nostra ignorància convençuts que hi estàvem capacitats, ens vam oferir a ajudar-lo fent classe de problemes.



Com a premi-càstig per la nostra gosadia se'ns va habilitar un despatx al soterrani del Laboratori de càlcul. Aquest embrió del futur grup i departament, (adaptant la frase de Groucho Marx) "partint de la més absoluta pobresa va assolir els més alts cims de la misèria" quan li foren assignats dos despatxos situats en un sostre de l'edifici històric de la Universitat de Barcelona. Aquests despatxos, de parets corbades, tenien la propietat de tenir més metres quadrats que cúbics. Allí en Carles compartia espai amb persones que encara d'alguna manera ens considerem hereus acadèmics seus,

juntament amb persones que posteriorment van seguir altres camins, com ara en Guillem Alonso, la

Carmen Ruiz-Rivas o la Maite Usón. Malgrat algunes limitacions d'infraestructura com les esmentades, eren temps de treball entusiasta i de moltes esperances. Un record especialment entranyable, que va contribuir a estrènyer els lligams d'amistat, és la nostra participació al II SINAPE, un congrés de probabilitats i estadística a Campinas, Brasil, viatge que vam aprofitar per visitar Rio de Janeiro, Sao Paulo, les cascades d'Iguaçú i també Buenos Aires. Com a curiositat, recordo que quan l'avió estava començant a maniobrar per aterrar a Buenos Aires, ens vam assabentar que un tal Adolfo Suárez havia estat nomenat president del govern espanyol.

Un pas molt important en la consolidació del grup al voltant d'en Carles es va produir els anys vuitanta, quan la Facultat de biologia es va traslladar al seu actual edifici a la Diagonal. Allí ens foren assignats uns espais més adequats i posteriorment es va poder constituir el Departament d'estadística de la Universitat de Barcelona, del qual en Carles en fou el primer director. Una fita important d'aquells anys, tant personal d'en Carles com del grup, va ser la l'organització de las I JORNADAS DE BIOESTADÍSTICA, amb participació de col·legues i grups de Catalunya i de tot l'Estat espanyol que posteriorment han tingut especial rellevància en la docència i la recerca estadística en biologia, medicina i ciències afins. Amb molts d'aquests grups hem col·laborat i ho seguim fent. Els anys noranta del segle passat van ser anys de consolidació del prestigi internacional d'en Carles. Una fita important per la seva part va ser l'organització de l'edició de Barcelona de la Seventh International Conference on Multivariate Analysis, que va consolidar la seva (i del departament) relació amb el Prof. C.R. Rao.



Lupe Gómez:

El Carles Cuadras va ser el meu professor d'Anàlisi de la Variància, assignatura optativa a 5è a la Llicenciatura de Matemàtiques. Recordo les seves classes a l'aula 6, al grup reduït d'alumnes que havíem decidit fer de l'estadística la nostra passió, el Paco Carmona i l'Olga Julià entre ells. Després d'uns quants anys, a la meua tornada del Estats Units, vaig començar a coincidir amb el Carles a les Conferències de Biometria, les espanyoles i les internacionals. Em venen a la memòria records de la International Biometric Conference d'Amsterdam, el 1996, a la qual en Carles hi va tenir una participació destacada, com a organitzador d'una sessió dedicada a "Statistics for Environmental Data". La seva vinculació a la Biometric Society ja venia d'abans, de fet des de l'inici de la formació com Spanish National Group. El grup esdevingué regió i Carles fou el primer president de la Sociedad Española de Biometria (anys 1995-1996), de la qual jo n'era vocal i la corresponsal de cara a la societat internacional.

BIOMETRIC BULLETIN
International Biometric Society • Internationale Biometrische Gesellschaft • Société Internationale de Biométrie
VOL. 14, N° 2 JUNE 1997

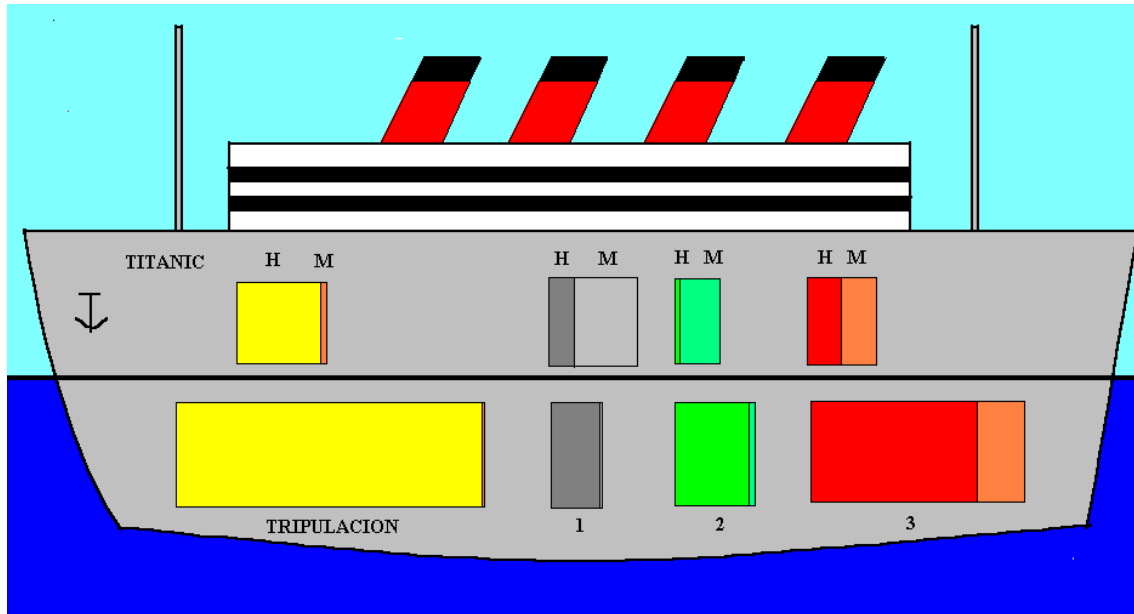
1995-1996
SPANISH

After the elections within the Spanish Region, the Council for the region has been formed as follows:


President: **Carles Cuadras** (carlesm@porthos.bio.ub.es)
Vice-President: **Emilio Carbonell** (ecarbo@via.es)
Secretary: **Fernando Lopez** (fernando@iata.csic.es)
Treasurer: **M. Dolores Sanchez**
Other members: **Juan Luis Chorro, Rosa Estarrelles, Guadalupe Gomez, Martin Rios**

The Council for the Spanish Region held a meeting in Valencia last February. The new President, Dr. Emilio Carbonell, moderated the Council meeting. The Treasurer, Dr. M. Dolores Sanchez, gave us the following data: we are so far 104 members including 11 students and 7 associates and the balance of the Region is slightly positive. The Secretary, Dr. Fernando Lopez, is in charge of the email list as well of maintaining the Web page. He informed that there are still many members to be included in the email list. He is also encouraging us to add our personal page link to the Web (<http://www.iata.csic.es/BSREsp/>). The past President, Prof. Carles Cuadras, informed of his achievements during his 2 years service. They include, among others, the bilateral agreement with the Statistical Institute of Catalunya. Under such cooperation the spanish journal QUESTIO has started a Biometry Section (see Biometric Bulletin Vol 13, N. 4 and all the members of the Spanish Region of the International Biometric Society that receive Biometrics will also receive QUESTIO as well as Biometrics). We sincerely thank Dr. Cuadras for his job. Dr. Carbonell set up his goals for his two year presidency. Among others include to get more memberships, mainly from the underrepresented regions in Spain, to facilitate information about the several issues of concern for our society, mainly through the Web page, and to organize courses and workshops under the auspices of the Spanish region. The VI Conferencia Española de Biometría will take place from September 21 through 24, 97 in Córdoba. For more information check the Region Web page and also <http://www.uco.es/~biometry>.
Lupe Gómez
Spanish Correspondent

Un altre record, potser un dels majors cims de l'estadística jocosa, correspon a la presentació que improvisà a la VIII Conferencia Española de Biometría de 2001, a Pamplona, per salvar l'atzucac en que ens trobàvem atesa la nul·la participació en allò que els organitzadors havíem denominat "El Reto Titánico", una invitació a analitzar unes dades de freqüències de supervivència i mortalitat derivades del naufragi del Titànic ocorregut el 1912. En aquella presentació es va introduir, en exclusiva mundial, el "nautigrama" (veieu gràfic adjunt).



Com ja s'ha apuntat anteriorment, els meus contactes amb el grup d'en Carles venen de lluny. De tota manera voldria ressaltar que els darrers anys la col·laboració ha esdevingut molt més estreta,

tant a nivell docent com a nivell de recerca. El GRBIO,  "Grup de Recerca en Bioestadística i Bioinformàtica" grup consolidat de la Generalitat de Catalunya, és hereu directe del grup consolidat "Grup d'estadística multivariant i computacional" del qual en Carles n'era l'IP.

El Carles em va donar també l'oportunitat de col·laborar, primer com a editora associada i després executiva, a la revista Qüestiió, de la que ell n'era l'Editor. Segur que el Carles recorda aquell estiu a casa seva amb en Pere Puig i en Michael Greenacre decidint que fer amb Qüestiió, animant-nos a donar un tomb a la mateixa i rient buscant noms i bon acrònims. Aquell dia d'estiu, a Tossa de Mar, va néixer la revista SORT.

Agraeixo al Carles tot el que ha fet per la Sociedad Española de Biometria, per les revistes Qüestiió-SORT i sobretot , per la confiança que ell ha tingut sempre amb mi, per incloure'm en algunes de les seves iniciatives , donant-me d'aquesta manera la possibilitat de conèixer-lo millor i de gaudir de moltíssimes xerrades sobre estadística i sobre "el bien y el mal".

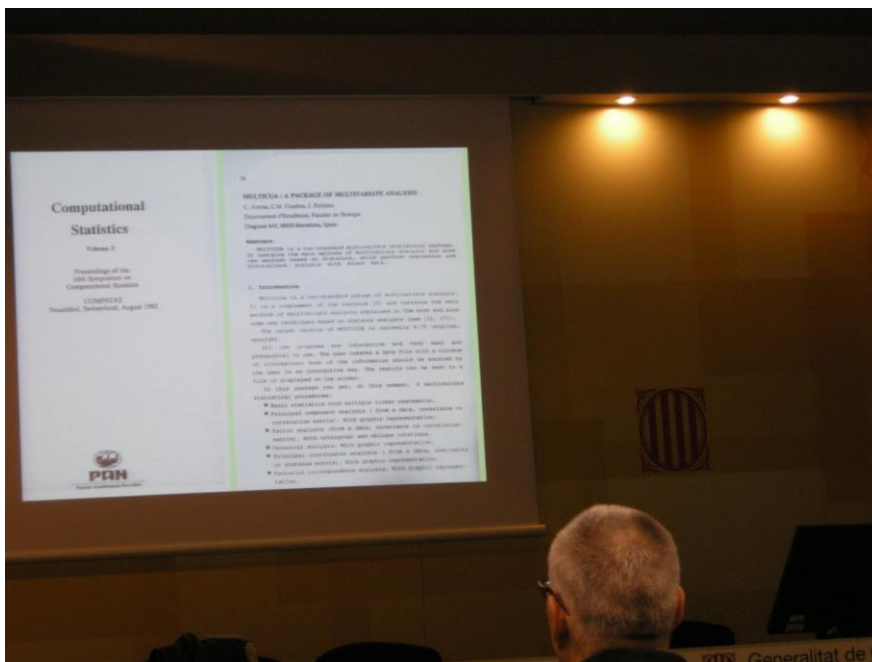
Jordi Ocaña i Lupe Gómez:

No som pas els únics, però sens dubte per nosaltres és un mestre i un amic.



Una forta abraçada

Conchita Arenas, University of Barcelona, Spain.



RECALLING SOME CONGRESSES

C. Arenas

Department of Genetics, Microbiology and Statistics. Statistics Section
University of Barcelona, Spain

Prof. Cuadras was who introduced me to the world of the congresses and allowed me to discover its usefulness. The congresses allow us to know specialists of topics we are interested in; we can talk with these specialists and to learn from them. In this way, reading a published paper is more human because we can know the personality of the author.



From left to right: M. Greenacre, J. M. Oller, C. M. Cuadras, C. R. Rao and his wife

Furthermore, the congresses have a cultural side, allowing us to know cities or places of interest. On the other hand, it is very important to attend communications presented in congresses as we can learn a lot from all of them.



Prof. Cuadras in Paris

I remember with special fondness few presentations made on behalf of the group led by Prof. Cuadras about the Multicua project. Multicua was a free analysis multivariate package. It contained standard methods as principal component or principal coordinate analyses, and distance-based methods, such as regression or discriminant analysis based on distances that are part of another of the lines of research of Prof. Cuadras. This package was very easy to use and it was used both as a teaching and research tool. Some of the multiple papers in which Multicua was used are: Jakimiee and Bartkowiak (1994); Pedrola-Monfort and Caujapé-Castells (1994); Fleck et al. (1996); Fernandez-Leborans (2000); Jaramillo-Correa et al. (2001); Oliver et al. (2002); Arribas and Carranza (2004) or Arribas et al. (2006) between others. This package was presented in the XIX Reunión de Estadística, Investigación Operativa e Informática which held in Segovia (Spain) in 1991; in the 10th Symposium on Computational Statistics (COMPSTAT) in 1992 (Neuchâtel, Switzerland) and as software demonstration, in the 7th International Conference of Multivariate Analysis held in 1992 (Barcelona, Spain).



Attending the 7th International Conference of Multivariate Analysis meeting.

The circle points Prof. Cuadras

It has been a pleasure for me to share many good moments in congresses with Prof. Cuadras.

REFERENCES

Arenas C; Cuadras CM; Fortiana J (1991). Multicua. XIX Reunión de Estadística, Investigación Operativa e Informática, Segovia (Spain).

Arenas C; Cuadras CM; Fortiana J (1992). Multicua: a package of multivariate analysis. 10th Symposium on Computational Statistics (COMPSTAT), Neuchâtel (Switzerland).

Arenas C; Cuadras CM; Fortiana J (1992). Multicua: a non-standard package for multivariate analysis. 7th International Conference of Multivariate Analysis, Barcelona, (Spain).

Arenas, C; Cuadras, CM; Fortiana, J. (1998). MULTICUA: Paquete no estándar de Análisis Multivariante, versión 0.77 ampliada. Publicacions del Departament d'Estadística (nueva colección), Barcelona (Spain).

Josep M^a Oller, University of Barcelona, Spain.



ALGUNOS RECUERDOS DE MIS COMIENZOS EN LA UB Y EL DR. CARLES M CUADRAS

Josep M. Oller (UB)

La primera referencia que tuve del profesor Carles M. Cuadras fue en septiembre 1974 (curso

1974-75) a través de sus libros de problemas de Cálculo de Probabilidades y Estadística, editados por el Laboratorio de cálculo, y recomendados por mi profesor de entonces, Guillermo Alonso. Dicho profesor me sugirió que asistiese de oyente al cursillo de doctorado, curso 1975-76, que impartía el profesor Cuadras. Así lo hice. Fue mi primer contacto con temas de Análisis Multivariante, entre otros el Análisis de Componentes Principales, el Análisis de Coordenadas Principales, el Análisis Factorial de Correspondencias, el Análisis Canónico de Poblaciones, la Correlación Canónica... El profesor Carles M. Cuadras impregnó el cursillo de abundantes aspectos geométricos que empezaron a inclinarme hacia este aspecto de la Estadística. El cursillo se realizaba de buena mañana en la biblioteca del departamento de Genética de la Universidad de Barcelona, en aquel entonces situado en la plaza Universidad y recuerdo algunos de sus asistentes: el propio profesor Guillermo Alonso, otros profesores de la sección de Bioestadística como Jordi Ocaña, Mercè Argemí, Carmen Ruiz-Rivas... y otros profesores de genética y ecología como Antoni Escarré, Eduard Petitpierre, Miguel de Renzi... Esta experiencia me acabó de convencer que eligiera posteriormente la asignatura optativa de Ampliación de Bioestadística, que cursé el curso 1977-78 e que impartía el profesor Cuadras.

Dicha asignatura cubría la teoría básica de los modelos lineales normales, con el Teorema Fundamental del Análisis de la Varianza, Teorema de Gauss-Markov y diversos métodos de Análisis Multivariante. Se hacía hincapié también en la distancia de Mahalanobis que fundamenta considerables aspectos del Análisis Multivariante. En aquel momento recuerdo que me llamó la atención la relación entre independencia estocástica y ortogonalidad y, en general, el papel que jugaba la geometría en los métodos considerados, todo ello sin duda influido por la labor docente del profesor Cuadras.

Al terminar la licenciatura en Biología, y después del servicio militar en Madrid (Abril 79-Mayo 80) lo primero que realicé fue acabar y leer la tesina, que aún tenía pendiente, en Septiembre de 1980 y cuyo director fue el Dr. Antoni Prevosti, sobre un tema de estadística aplicada al estudio de la eficacia biológica. Seguidamente me incorporo como profesor ayudante al "grupo" de Bioestadística dentro del departamento de Genética, curso 80-81. Es allí donde conozco más estrechamente al profesor Carles M. Cuadras, también ya como director de tesis, sugiriéndome, en primer lugar, que realizara una tesis netamente de Biometría, donde debía profundizar en la definición y caracterizaciones multivariantes de los conceptos de tamaño y forma. Bajo su dirección empecé a estudiar artículos clásicos como:

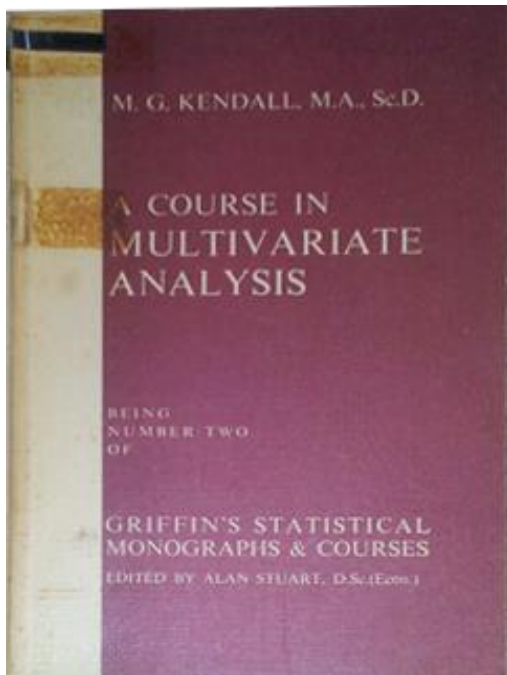
Burnaby, T.P. (1966) Growth-invariant discriminant functions and generalized distances, *Biometrics*, 22, 96-110.

Blackith, R. E. (1960) A Synthesis of Multivariate techniques to distinguish patterns of growth in Grasshoppers, *Biometrics*, 16, 28-40.

Sin embargo influido lo que había aprendido con Carles M. Cuadras, me atraía el papel de la geometría en el estudio de la Ciencia, especialmente en Física y seguía pensando en distancias naturales entre objetos estadísticos...

Un día, en una charla informal, Carles M. Cuadras me hizo una sugerencia de una referencia clave donde se indicaba todo un procedimiento de construcción de distancias entre poblaciones estadísticas correspondientes a un cierto modelo paramétrico. La referencia fue el siguiente libro de Kendall:

Kendall, M.G. (1957) *A Course in Multivariate Analysis*. Griffin's Statistical Monographs & Courses. C. Griffin and Co. Ltd. London.



En dicho libro se citaba un trabajo de C.R. Rao de 1949, *On the distance between two populations*, publicado en *Sankhya*, 9, 246-248. (Obsérvese: no el del propio Rao en 1945 sobre dicho tema). En este trabajo se exponía un procedimiento que permitía generalizar la distancia de Mahalanobis a poblaciones estadísticas paramétricas con algunas condiciones básicas de regularidad. El procedimiento descrito, como el artículo original de Mahalanobis (1936), utilizaba herramientas de análisis tensorial y geometría diferencial, que ya me habían llamado la

atención al tratar de comprender diversos temas de Física, como la Relatividad General. Dicha sugerencia me indicaba el camino de la tesis, centrado en distancias estadísticas y alejándome de la propuesta inicial.

El primer paso fue conseguir el trabajo de Rao (1949) pues en la Biblioteca de la Universidad de Barcelona la suscripción a *Sankhya* fue más tardía. Pedí una copia del trabajo y tras varios días me la enviaron un negativo ¡en microfilm! y que requería una máquina especial para leerlo en la biblioteca. Dicho microfilm lo hice positivar, copia positivada que aún conservo como recuerdo de una época cercana, donde la información no estaba tan al alcance de todos, como hoy en día lo es. Luego siguieron cálculos, lectura, revisiones, consejos, correcciones... y más cálculos, y vuelta a empezar... hasta la finalización de la tesis, donde se calculaba la distancia de Rao correspondientes a varios modelos estadísticos y se discutían algunas aplicaciones. La memoria se completó el 25 de noviembre de 1982 y se leyó el 23 de febrero de 1983, dos años después de *aquel* 23F. Destacaré de aquella memoria un artículo conjunto con el profesor C.M. Cuadras (1985) del que me siento particularmente contento:

Oller, J.M. and Cuadras, C.M. (1985) Rao's distance for negative multinomial distributions. *Sankhya* 47A, 75-83.

De dicho artículo destacaré la distancia entre distribuciones multinomiales negativas remarcada:

The integration constants will be chosen in such a way that they accomplish (2.18), and satisfy the initial conditions, that is, for $s = 0$ the geodesic departs from a point A , with coordinates (a^1, \dots, a^n) and for a positive value $s = d$, reaches the point B , (b^1, \dots, b^n) . After some laborious calculations we obtain that the geodesic distance between A and B is given by

$$d = 2\sqrt{r} \cosh^{-1} \left(\frac{1 - \sum_{j=1}^n \sqrt{a^j b^j}}{\sqrt{a^{n+1} b^{n+1}}} \right) \quad \dots \quad (2.29)$$

Mucho ha llovido desde entonces: clases, trabajos, la faena del día a día, la primera tesis que dirigí (Miquel Calvo)...

Es justo recordar hoy que muchas de estas actividades y resultados tuvieron su origen en la curiosidad que suscitó y la ayuda que me proporcionó el profesor Carles M. Cuadras y esta charla pretende agradecerlo.

No es ni mucho menos una exageración decir que Carles M. Cuadras, especialmente por su carácter pionero en introducir multitud de métodos multivariantes, ha realizado una contribución capital a la Estadística en Cataluña y en el resto de España. Recordemos, entre otros, su libro *Métodos de Análisis Multivariante*, el primer libro en castellano sobre dicha temática así como sus numerosos artículos de investigación. Por todo ello:

¡Gracias Profesor Cuadras!

Martín Ríos, University of Barcelona, Spain.



GRACIAS PROFESOR CUADRAS

Durante los estudios de primer ciclo de la licenciatura de medicina en la que entonces se llamaba Universidad de Santander hoy de Cantabria, me preguntaba que estadística de buen nivel matemático, se utilizaba en la investigación médica. Fue en el verano de 1979 cuando tuve la

primera referencia del libro de “Problemas de Probabilidad y Estadística” del profesor Carlos Cuadras, editado por EUNIBAR y que era seguido por los alumnos de la licenciatura de Biología en la asignatura de Bioestadística. Por ese motivo me trasladé a vivir a Barcelona e integrarme en el equipo del entonces profesor agregado Dr. Cuadras y compuesto por los profesores J. Ocaña, J. M. Oller, P. Sánchez, F. Carmona, M. T. Usón, M. C. Ruiz de Villa, A. Arcas, M. Salicrú, J. Fortiana, M. Argemí, M. L. Sanz, X. Viader, L. Serra (Genética) y C. Ruiz-Rivas (U.A.M.).

En los primeros años aprendí sobre todo la Teoría de los Modelos Lineales y Análisis de datos Multivariantes gracias a la publicación por parte del profesor Cuadras del libro “Métodos de Análisis Multivariante” que ha sido durante años el texto de estadística más citado en español.

De su enseñanza aprendí el papel que jugaba la geometría en la estadística y decidimos distanciar objetos estadísticos (poblaciones e individuos). Este trabajo dio como fruto la elaboración de mi tesis doctoral “Métricas entre Modelos Lineales y su aplicación al tratamiento de datos en Medicina” defendida en la Facultad de Medicina de la U.B. en Octubre de 1985. .

Como resultados más importantes de la tesis podemos resaltar:

La obtención de distancias entre modelos lineales univariantes con igual y distinta varianza, distancia entre modelos multivariantes y aplicaciones a la medicina como una clasificación de los TTOG realizados a 172 niños donde se medía la glucemia y la insulínemia, lo que me permitió aplicarlo al caso multivariante y un algoritmo bayesiano para diagnosticar los TTOG.

Por el carácter humano y científico del profesor Carlos María Cuadras deseo expresarle desde aquí mi más sincero agradecimiento.

Martín Ríos

Daniel Cuadras, Fundació Sant Joan de Déu, Spain.

Dr. Cuadras junior



Josep Fortiana, University of Barcelona, Spain.



Josep Maria Duran, University of Barcelona, Spain.

RECORDANT MOLTS BONS MOMENTS

Bon dia Carles.

Vagi per endavant que em sap molt de greu no ser personalment en aquest acte. Malauradament a l'hora programada perquè els "tutorats" teus diguin unes paraules hi ha una sortida d'àrea dels 1rs de batxillerat de l'escola on treballo i, com sóc tutor d'un grup, és imprescindible la meua assistència. És el que té ser professor/tutor de secundària. No s'entendria que no estès amb el meu alumnat en una sortida i més essent la primera des que els conec. Des de que sóc el seu tutor.

Això em dona peu a dir-te una cosa que havia comentat en el seu moment amb els companys de l'escola. Els hi deia, el meu tutor de Tesi, en Carles M. Cuadras, seria un fantàstic professor de secundària. És de les millors coses que et puc dir essent, com sóc, professor des de fa més de 40 anys.

Un segon assumpte que et vull comentar. Ara respecte la tesi i el temps que vaig tenir l'honor que m'atenguessis; he de dir que em vas explicar les coses amb una proximitat, una dedicació, un interès i una claredat que he agraït des d'aleshores.

Una tercera qüestió. Em vas acollir com alumne tot i tenir ja una certa edat i estar una mica, o molt, desconnectat de la vida acadèmica (cursos de doctorat a part) i, diguem-ho clarament, no gaire brillant. Gràcies un altre cop.

Un últim comentari i no menys important. Gràcies, ara, per fer-me partícip de la creació de "Report" un dels teus llibres, diguem, de ficció. Recordaré sempre els moments del carrer Manacor en els quals et comentava el que anava llegint, moments tant interessants com els dedicats a les desigualtats en regressió i anàlisi multivariant. Va ser una experiència única que he tingut el plaer de viure.

Gràcies, finalment, tot i que fa molt de temps que no ens veiem, per la teva amistat.

Una abraçada, la qual espero donar-te personalment aquesta tarda.

Josep Maria Duran Rúbies

Barcelona, dijous 20 d'octubre de 2016

Joan Manel Batista, ESADE BS. Universitat Ram3n Llull, Spain.



CARLES CUADRAS

Hola Carles, un curs fora no em permet acompanyar-vos. Tinc l'orgull de ser el teu segon doctorand despr3s de Carmen Ruiz Villa (la de Madrid). A les hores tu, amb ajudants de luxe -Alonso i Ocaña, portaves el centre de c3lcul de la UB. No va ser f3cil, i no sols per que a finals dels 70 la "transici3" (des de el PSUC) va ser dura, si no perqu3, el que un PNN de enginyers que estudiava psicologia escollis un matem3tic per dirigir-li la tesis, no va agradar a algun catedr3tic. Ho recordes?

Jo recordo les esperes en els patis de la UB per corregir amb tu els minsos avanços que feia, fins que em vas descobrir Essex (on em vaig passar 7 estius). Per3, del cam3 que em vas ajudar a rec3rrer primer tu, i despr3s Willem Saris vaig aprendre a fer recerca, i fixat, es amb el que encara ara em guanyo la vida.

Alg3 va dir fa temps, que l'exemple no 3s el m3s important per influir en els altres, 3s l'3nic, l'3nic que conta. I pot estar segur, que ha estat el teu exemple el que ens ha influ3t a tants, i a tots els participen en aquest homenatge.

Nom3s desitjaria deixar en els meus doctorands un record semblant al que has deixat tu en els teus.

Gr3cies Carles.

Purificación Galindo, University of Salamanca, Spain.





INVESTIGACIÓN DE LOS HIJOS NIETOS Y BISNIETOS DE CARLOS CUADRAS FORMADOS EN SALAMANCA: PASADO, PRESENTE Y FUTURO

Purificación Galindo Villardón. Dpto. Estadística. USAL. España. pgalindo@usal.es

RESUMEN



Hace ahora 30 años, se presentaba la primer tesis doctoral dirigida por el Profesor Cuadras, en la Universidad de Salamanca. La tesis, defendida por **Purificación Galindo, hija científica**, presentaba una alternativa de representación simultánea de datos multidimensionales, llamada **HJ-Biplot**, que fue publicada en la revista *Questiío* en **1986**, la cual ha sido aplicada en muchos campos de la ciencia (Medicina, Biología, Ingeniería Civil, Ciencias Ambientales, Sociología, Ciencia Política, Neuropsicología, Limnología, Gestión Hospitalaria, Inferencia Ecológica, Biotecnología, Bibliometría, Sostenibilidad, Salud Ambiental y Toxicología,... y un larguísimo etc.) y ha recibido una alto número de citas, tanto en la WoS como en Google Scholar Citation. En *Questiío* ha sido el segundo trabajo más citado, superado solo por un trabajo de Rao.

Con esta tesis doctoral, el Prof. CUADRAS abre una línea de investigación sobre los métodos **BIPLOT**, que se ha mantenido activa y ha producido importantes contribuciones, durante todos estos años. En la tesis doctoral de [José Luis Vicente-Villardón](#), primer *nieto científico*, 1992, se proponía el **HJ-Biplot generalizado** en el que podían utilizarse métricas distintas de la identidad, tanto para las filas como para las columnas, de una matriz de datos. El método permitía, por ejemplo, realizar Biplots ponderados. Recibió Premio Extraordinario (**PE**). Como caso particular de los Biplots ponderados, se describieron los Biplots Canónicos (**MANOVA Biplot**) que después fueron desarrollados con mayor detalle y extendidos a diseños experimentales con más de una vía por [Rafael Amaro](#), *bisnieto*, en 2001. Estos Biplots se han aplicado en más de 20 artículos en el entorno del Departamento, publicados en revistas indexadas en el Journal Citation Report (JCR) y se siguen utilizando en aplicaciones. Es una línea de investigación abierta, especialmente la utilización en conjuntos de datos con más variables que individuos. [Santiago Vicente-Tavera](#), *nieto*, también en 1992, presentó el **Método de Inercia** que combina los métodos Biplot y el Análisis de Clúster sobre coordenadas Biplot. El método ha sido utilizado en muchos trabajos entre los cuales merece especial atención el realizado por [Carmen Patino](#), *nieta*, 2008 para realizar la caracterización multivariante de la realidad socio-económica de la mujer salmantina con empleo irregular, que recibió premio a la excelencia investigadora.

[Valter Vairinhos](#), *nieto*, en 2003 desarrolló un sistema para Minería de Datos basado en los métodos BILOT, llamado **BiplotsPMD** y [Sergio Hernández](#), *nieta*, en 2005 desarrolló versiones robustas de los Biplots clásicos (GH, JK y HJ Biplots), **BILOTs Robustos**, y un software específico para su aplicación.

[Ana Nieto](#), *nieta*, en 2015 presenta una versión inferencial de los métodos BILOT basada en remuestreo Bootstrap y lo extiende a tablas de dos y tres vías y desarrolla, además, un nuevo método llamado **CLUSTERING DISJOINT HJ-BILOT** que extiende las ideas de Amaro, y aborda el doble problema de clasificación y reducción de la dimensión, proponiendo soluciones con componentes principales disjuntas y recogiendo la estructura en grupos. Desarrolla sendos paquetes en entorno R, el **biplotbootGUI-R** y el **MultiBiplotGui-R**, para poder llevar a la práctica estas propuestas.

[Zulaima Osuna](#), *nieta*, en 2006 presenta contribuciones al análisis de datos textuales basadas en métodos Biplot.

Los Biplots con ponderaciones diferentes para las filas y/o las columnas pueden extenderse para

utilizar ponderaciones distintas para cada uno de los elementos de la matriz mediante la utilización de algoritmos de mínimos cuadrados alternados. Esto permite la extensión de los métodos a datos con distribuciones de la familia exponencial, [Antonio Blázquez, nieta, 1998 \(PE\)](#). [Olesia Cárdenas, nieta, 2000](#), desarrolló los **Biplots con información externa**, basado en modelos lineales generalizados. En esta línea [Maura Vásquez, nieta, en 1995](#) presentó interesantes aportaciones en el tratamiento de datos faltantes.

Cuando se trata de tablas de doble entrada, en las que todos los elementos de la matriz proceden de una distribución del mismo tipo, estos métodos son particularmente útiles para el diagnóstico de modelos más formales ([Marisol Díaz-Leno, nieta, 1995](#), y [Rosa Sepúlveda, nieta, 2004](#)). El diagnóstico se ha empleado, por ejemplo, para la detección de dependencias condicionales en el modelo de clases latentes (Rosa Sepúlveda).

En tres tesis doctorales se estudian los modelos de Rasgos Latentes y su posible representación mediante Biplots ([Miguel Yáñez, bisnieta, 2003](#)), los modelos de la TRI para datos binarios ([Hermes Viloría, bisnieta, 2004](#)) y los modelos de TRI aplicados en Calidad de Vida ([Mercedes Sánchez-Barba, bisnieta, 2008](#)), todos ellos relacionados con datos categóricos, tanto binarios como nominales u ordinales.

Los algoritmos alternados, junto con los métodos para datos categóricos, se han empleado también para la modelización de tablas con datos binarios (**Biplot Logístico**). El algoritmo para datos binarios se extendió a casos en los que el número de variables era elevado proponiendo un algoritmo externo basado en las Coordenadas Principales ([Jhonny Demey, nieta, 2008](#)). El método es similar, en cierto modo, a los Análisis de Regresión basados en distancias que están entre las líneas de investigación del Prof. Cuadras. Los Biplots Logísticos se han aplicado en diversos campos, en particular, en el ámbito de la innovación y la sostenibilidad, en revistas de alto impacto.

Recientemente se ha extendido a datos nominales y ordinales, [Julio César Hernández- Sánchez, bisnieta, 2016](#). Desarrolla, además, softwares específicos para su aplicación, **NominalLogisticBiplot-R** y **OrdinalLogisticBiplot-R**. Esta sigue siendo una línea de investigación abierta.

Desde otro punto de vista, los BIPLoTs para datos continuos se han generalizado al caso en el que se dispone simultáneamente de varias matrices de datos:

En el estudio de las relaciones entre abundancia de especies y variables ambientales mediante la utilización del método Unfolding. En este caso se propone un algoritmo de Análisis Directo del

Gradiente, alternativo al CCA de Ter Braak, basado en Unfolding (M^a José Fernández-Gómez, *nieta*, 1995) que integra directamente la información ambiental en el cálculo de las coordenadas de las especies. El resultado es un TriPlot en el que se combinan la interpretación basada en distancias, procedentes de los modelos de preferencia, con la interpretación basada en productos escalares, procedente de los biplots. Una alternativa asimétrica al CCA (**CNSCA**) fue propuesta por Priscila Willens, *nieta*, 1995. En 2002, Sonia Salvo, *nieta*, propone una alternativa a las propuestas de la escuela italiana para el tratamiento de tablas de correspondencias asimétricas, de tres vías, basada en el modelo TUCKER3.

En el estudio de varias matrices de datos, cuando las filas están divididas en grupos, se propuso el que denominamos STATIS Canónico (**CANOSTATIS**), Amparo Vallejo, *bisnieta*, 2004, que permite la representación consenso de los grupos en diferentes momentos del tiempo, junto con las variables responsables de la separación.

En el estudio de varias matrices de datos con variables comunes se desarrolla un marco común a varias técnicas, basado en la obtención de subespacios consenso, en los que proyectar todas las matrices, con el fin de obtener un Biplot Consenso para todas ellas, en el cual las coordenadas de las variables sean comunes. (Luz Mari Pinzón, *bisnieta*, 2012). Esto puede considerarse como una extensión a las técnicas de la escuela francesa y una extensión del **META-BIPILOT** propuesto anteriormente por Jesús Martín Rodríguez, *nieta*, en 1996, y de las representaciones Biplot para el Análisis Triádico (Luiz Basso, *bisnieta*, 2006). En 2005, Ángela Cortés, *bisnieta*, presenta contribuciones al Análisis de Tablas de Tres Vías Restringido, **CNSCA de tres vías**. Nora Baccalá, *nieta*, en 2002 presenta el **MULTIBIPILOT**; es decir la extensión del Análisis Factorial Múltiple al caso Biplot de datos continuos de tres vías. Javier Martín-Valejo, *nieta*, en 1995 introduce los Meta-BIPILOT en el contexto del Meta-análisis.

Puri Vicente-Galindo, *nieta*, en 2003 (**PE**), propone los métodos STATIS y los Meta-BIPILOT como herramienta de diagnóstico del tipo de cambio (Alfa, Beta y Gamma), en estudios de Calidad de Vida Relacionada con la Salud, para detectar Response Shift. En 2010, Sara Nunes, *nieta*, continúa trabajando en modelos para evaluar la Calidad de Vida percibida y compara los resultados obtenidos con estos métodos y los obtenidos con el modelo de Zwinderman y las curvas de crecimiento, en pacientes oncológicos. En 2015, Francisco Perdomo, *nieta*, presenta, en esta misma línea, el trabajo Evaluación del Cambio en la Calidad de Vida de pacientes Miopes operados con Cirugía de Láser Excimer corneal, que es el primer trabajo en el mundo inspeccionando Response

Shift, en presencia de eventos positivos, con estos métodos.

El estudio de la interacción de orden superior en modelos bilineales (trilineales), especialmente los utilizados en la detección de la interacción genotipo-ambiente con un tercer factor, por ejemplo el tiempo, fueron realizados por [Mario Varela, bisnieto](#), en 2002. En 2015, [Elisa Frutos, nieta](#), desarrolla el software **GGEbiplotGui** el cual está siendo utilizado por universidades y centros de investigación de todo el mundo.

En 2006, [Roxana Ríos, nieta](#), presenta aportaciones al análisis dinámico de tablas de tres vías, y [Susana Mendes, nieta, \(PE\)](#) presenta en 2006 el **TUCKERCo**, un método multivariante para evaluar patrones de estabilidad y cambio, desde una perspectiva Biplot, que combina un análisis de Coinercia con el Modelo TUCKER3.

En 2015, [Jaime Egido, nieta](#), propone el **BIPLOT DINÁMICO** para tablas de tres vías, que presenta la ventaja de que en lugar de tomar como referencia una matriz consenso permite elegir cualquiera de las matrices individuales como referencia y analizar la evolución del fenómeno en relación a ésta. Desarrolla el paquete **dynbiplotGui-R** para su aplicación.

La última tesis defendida en 2016, [Miguel Rodríguez-Rosa, nieta, \(PE\)](#) extiende los trabajos de Ríos y Mendes (op. cit.) y presenta el **CoTUCKER3**, con una interesante aplicación al análisis de la sostenibilidad internacional.

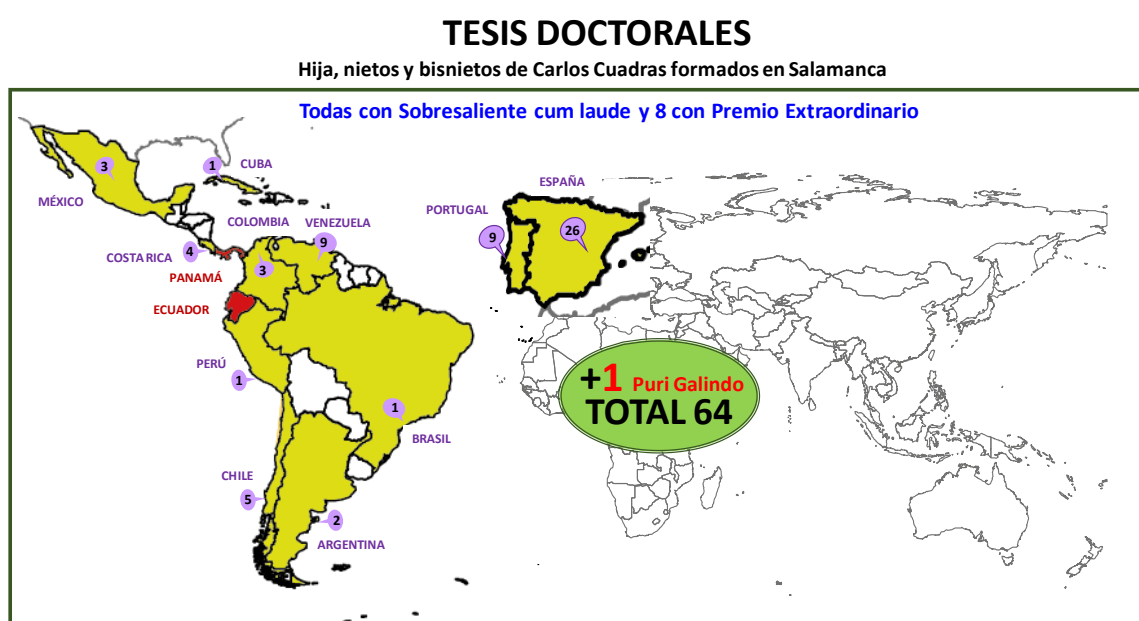
En 2016, [Javier Delgado, nieta](#), presenta Contribuciones basadas en el análisis BILOT al diseño y gestión de redes de telecomunicación.

En 2015, [Elisa Frutos, nieta](#), presenta el trabajo Análisis de Datos acoplados: **Modelo T3-PCA** para el análisis de un bloque de datos de tres vías y un bloque de datos de dos vías, ambos de tipo continuo, que tienen un modo común. Se construye un modelo que conste de dos submodelos, un modelo Tucker3 para el bloque de datos de tres vías y un modelo PCA para el bloque de dos vías. Los submodelos se relacionan imponiendo la restricción de que la matriz de componentes del modo común ha de ser igual en ambos submodelos. Se propone, además, un algoritmo para el cálculo de los parámetros del modelo, un método para el cálculo del rango y rotaciones que mejoran la interpretabilidad. Se desarrolló el software necesario para su aplicabilidad.

Hay otras muchas tesis que no se detallan porque no están dentro de la línea BILOT que es la línea abierta por el Prof. Cuadras hace 30 años, en el Departamento de Estadística de la USAL, pero también ellos se sienten nietos o bisnietos porque también han recibido sus enseñanzas, en directo

o a través de sus artículos y entre las que cabe destacar las que representan importantes contribuciones dentro de los métodos supervisados de Data Mining, otra de las líneas fuertes del Departamento de Estadística: [Guillermo Ramírez, nieta, 1995](#), [Carmelo Ávila, nieta, 1996](#), [Ana Dorado, nieta \(PE\), 1998](#), [Claudio Castro, nieta, 2005](#).

En total, en estos 30 años, se han realizado, **64+1 tesis doctorales**, todas ellas han recibido la calificación de Sobresaliente cum laude (o similar) y **8 han recibido Premio Extraordinario**. Los nietos y bisnietos del Dr. Cuadras, han llegado desde **11 países** y ahora desarrollan allí su actividad profesional. Ver mapa de distribución más abajo.



Países de procedencia de los nietos y bisnietos de Carlos Cuadras

En la actualidad, el Programa de Doctorado en Estadística Multivariante Aplicada tiene inscritos muchos estudiantes de diferentes países pero los grupos más numerosos han llegado de la ESPO de Guayaquil, Ecuador; del Centro de Investigación en Estadística Multivariante Aplicada (CIEMA) de Colima, México y de la Universidad de Panamá. Los trabajos de investigación continúan en las líneas anteriores, estudiando las técnicas factoriales reformuladas como técnicas de regresión penalizada que facilitan la interpretabilidad introduciendo la penalización Elastic net, o Lasso, en el problema de optimización (SPARSE PCA) y/o la descomposición CUR. Se centran en análisis estadístico de BIG DATA, Data Mining, Machine Learning y Modelos Bayesianos en Meta análisis, entre otros.



Profesores de la ESPOL de Guayaquil, Ecuador, en el doctorado Estadística Multivariante Aplicada del Departamento de Estadística de la USAL, España.



Investigadores del CIEMA, México, en el doctorado Estadística Multivariante Aplicada del Departamento de Estadística de la USAL, España.



Investigadores Colombianos, Panameños, Ecuatorianos, Españoles y Mexicanos, en el doctorado Estadística Multivariante Aplicada del Departamento de Estadística de la USAL, España.

Gracias Profesor Cuadras, sin ti nada de esto hubiera sucedido



Karina Gibert, Universitat Politècnica de Catalunya – BarcelonaTECH, Spain.



EN HONOR DEL DR. CARLES CUADRAS AVELLANA

amb tot l'afecte i admiració

Karina Gibert

Dep. Estadística i Investigació Operativa

Knowledge Engineering and Machine Learning group

Universitat Politècnica de Catalunya-BarcelonaTECH

Barcelona, 21 octubre 2016

La primera vegada que vaig sentir parlar del Dr. Carles Cuadras, jo era una nena de 17 anys que estudiava a l'institut més antic d'Espanya, el Ramon Muntaner de Figueres. Va ser a classe de matemàtiques. El meu professor ens explicava que tenia un germà molt important a la Universitat, a Barcelona, que li deia que la Universitat no tenia res a veure amb l'institut i que ens preparéssim per estudiar molt més i treure notes molt justetes...

Lavors jo sabia que volia estudiar informàtica, però no tenia ni idea que acabaria fent carrera a la universitat, ni molt menys, que m'acabaria dedicant a temes estadístics.

Quan, a punt d'acabar la carrera, em van fitjar al Departament d'Estadística i Investigació Operativa de la Facultat d'Informàtica de Barcelona, a la Universitat Politècnica de Catalunya, vaig començar-me a pensar de fer una tesi.... No sé massa ben bé com, em vaig trobar en un tema de clustering per dades heterogènies que em donava molt de mals de cap. Algú del departament em va dir que no teníem experts en això allí, i em va tornar a parlar del Dr. Cuadras, devia ser el 1991. Així que, un dia, em vaig decidir a anar-los a veure... i ell em va enviar al sòtan de la Facultat de Biologia, a la biblioteca, a buscar un paper del Biometrika d'un tal Gower.... i fou la primera vegada que vaig consultar un paper jo... i no em vaig ni atrevir a explicar-li que jo havia estat alumna del seu germà a Figueres, la seva ciutat natal. (quan en la seva lliçó inaugural dels Workshop, brillant per cert, el Dr. Cuadras va posar en un mapa factorial imaginari els científics que coneix, i em va posar vora el Gower, em va resultar tremendament entranyable!).

No va ser fins el 2006, quinze anys més tard, que en una SEIO a Tenerife, me'l vaig trobar assegut al meu costat a la terrassa de l'hotel. Era el vespre en que jo acabava d'arribar i necessitava menjar una mica. El tenia ben bé al costat i vam establir conversa, però de primer moment no el vaig reconèixer. Quan, després d'una bona estona de xerrera, ens vam presentar i em vaig adonar que parlava amb el Dr. Cuadras em semblava impossible tenir-lo allí mateix. Per mi era com un mite, m'havia passat mitja tesi donant voltes a la dissimilitud de Gower com a conseqüència d'aquella visita al seu despatx de Biologia, i mitja vida acadèmica explicant probabilitat i estadística a estadístics i informàtics amb el seu llibre!

I allí el tenia, amb aquest aire de despistat amable que només els savis tenen, explicant-me, mentre discutíem de multivariant, de còpules, de distàncies mixtes i de docència, que estava escrivint una novel·la, i que li agradava fer òpera amb titelles per a nens petits.... que tenia al cap fundar un grup d'interès en temes de multivariant lligat a la SEIO.... i mentrestant, jo agafava consciència plena que estava davant d'un dels grans mestres de l'estadística d'aquest país, en el sentit més profund de la paraula MESTRE. Font inesgotable d'il·luminació, cada conversa amb ell és un aprenentatge sense fi.... Un plaer... un gran plaer; conversar amb el Dr. Cuadras és sempre estimulante, inspirador i enormement plaent, no puc dir altra cosa.

Sovint he comprovat com n'és de famós... sense més, per Setmana Santa vam visitar un col·lega de la Universitat de Bath. Té 40 anys i és estadístic per la Universitat de Valladolid. Vaig intercanviar uns whats apps amb el Carles mentre era allí i sense voler en vaig dir el nom... M'obre uns ulls com dues taronges i em diu.... *"Cuadras?... el Cuadras del libro? "*, *"Sí, sí"contesto jo. "Peroooo, ... tu*

le conoces personalmente?...” i en aquell precís instant vaig reviure la mateixa sensació de mite que jo havia tingut a Tenerife.

Després de Tenerife, hem mantingut contacte regular amb formes i motius diversos: altres SEIOs, l'AMyC, altres congressos, actes culturals, tertúlies, seminaris, tribunals de tesi, les seves novel·les, hem parlat de majorettes i de música comparada,... el viatge a la India per l'aniversari del Rao!!!! inoblidable experiència aquella... i tantes altres tardes d'intercanvi personal i científic que no s'acaben mai.

Jo no he tingut el privilegi de ser alumna seva, però el segueixo amb afició, amb l'interès dels més fidels deixebles, i no em canso de meravellar-me de la seva incansable inquietud i qualitat humana i científica. I com heu pogut veure pel treball que acabo de presentar, aquella conversa dels anys 90, encara em dona fil de recerca en l'actualitat.

Carles, és un gran honor poder-te comptar entre les meves més apreciades amistats i poder compartir amb tu el nostre esperit empordanès, sempre vivent, i el teu gran saber.

Gràcies pel teu llegat

HONORING DR. CARLES CUADRAS AVELLANA

with all my sympathies and admiration

Karina Gibert

Department of Statistics and Operations Research

Knowledge Engineering and Machine Learning group

Universitat Politècnica de Catalunya-BarcelonaTech

Barcelona, october, 21st, 2016

The first time heard about Dr. Carles Cuadras, I was a teenager, 17 years old, and I was studying at the oldest Spanish high school, the Ramon Muntaner Institute, in the city of Figueres (Girona province). It was in my math class. My math teacher was telling us that he had a brother, very important the University in Barcelona, who explained to him that University is very different from high schools and we need to be ready to study much more at University and still get worse marks...

By then, I had a clear idea that I would enrol for a Computer Science degree, but I didn't know that I would follow an academic career, and I even could imagine that I will be linked to statistical topics.

About ending my Computer Science Degree, I was caught by the Department of Statistics and Operations Research from the Facultat d'Informàtica de Barcelona, at Universitat Politècnica de Catalunya, and I started to think about making a PhD thesis.... I do not know very well how I ended up with a problem related with clustering for heterogeneous data which was very difficult for me. Someone in my department told me that we did not have experts in that topic there, and addressed me to Dr. Cuadras; it was around 1991. Hence, I decided to visit him.... and he sent me to the basements of the Biology Faculty, to the library, to find a paper from *Biometrika* written by some Gower... and it was the first time that I was consulting a paper ... and I didn't have the courage to explain him that I had been a pupil of his brother in Figueres, his native city (when in the Workshops' opening lecture, certainly brilliant, Dr. Cuadras shown an imaginary factorial map with all scientists he knows, and he placed me near Gower, I was so touched!)

I did not meet him again till 2006, fifteen years later, in a SEIO in Tenerife. It was the evening of my arrival and I sit at the hotel terrace to eat something. He was near me and we started to talk, but initially, I could not recognize him. After a while, we introduced each other, and I realized I was talking to Dr Cuadras..... it seemed impossible to have him in front of me. He had become like a myth for me... I spent half of my PhD thesis working around Gower's similarity coefficient as a consequence of my visit to his desk in Biology Faculty, and half of my academic life teaching probability and statistics to computer scientists and statisticians using his book!

And he was there.... with this lovely absent-minded look that only wisemen have, explaining to me, while we discussed about multivariate analysis, copulas, mixed distances and teaching, that he was writing a novel, that he liked making opera with puppets for children... that he intended to found an interest group on multivariate topics linked to the SEIO... and I was becoming aware that I was in front of one of the grandmasters of Statistics in our country, in the deepest sense of the word GRANDMASTER. Endless source of illumination, each conversation with him is a never-ending learning... A pleasure... a great pleasure; talking to Dr. Cuadras is always stimulating, inspiring and greatly pleasant, I cannot say otherwise.

I could often verify how famous he is... last Easter, we visited a colleague of the University of Bath. He is 40 years old and is a statistician from the University of Valladolid. Being there, I was exchanging some WhatsApps with Carles and I mentioned his name. His jaw dropped and he said: "Cuadras?..."

the author of the book? “, “Yes, yes” I answered. “Buuuutttt, ... do you know him in person?....” and in that precise moment I evoked the same myth feeling that I experienced in Tenerife.

After Tenerife, we kept regular contact under different forms by several reasons: other SEIOs, l’AMyC, other conferences, cultural events, conversations, seminars, thesis commissions, his novels, talking about majorettes and compared music... the journey to India for Rao’s birthday!!! unforgettable experience... and so many other evenings of personal and scientific exchange that never end.

I had no the privilege to be his student, but I follow him with devotion, with the interest of the most loyal disciples, and I never stop to be amazed of his tireless curiosity as well as his human and scientific quality. And, as you may have seen by the work I presented in the workshop, that conversation from the 90s, still inspires me research nowadays.

Carles, is a great honor to be able to count you among my most appreciated friends and to be allowed to share with you our “empordanès” spirit, always alive, and your great wisdom.

Thanks for your legacy

José M^a Gutiérrez, University of Cádiz, Spain.

HOMENAJE AL QUERIDO PROFESOR CUADRAS

La verdad es que me entristece no poder estar en este instante acompañándote en tu homenaje. Gran parte ha sido mi culpa por no haber leído a tiempo la convocatoria del Workshop. Debido a la premura de tiempo me he encontrado sin un hotel, asequible a mi bolsillo, y sin billete de avión. A pesar de todo pensaba tomar algún tren con los transbordos oportunos y plantarme en Barcelona. Pero a todo ello se ha sumado un pequeño inconveniente de tipo biológico, debido al estrés del jubilado. En estos momentos estoy imbuido en tantas actividades que no adivino como he tenido tiempo, en mi vida laboral, para ir a trabajar. Espero que tomes nota y no te ocurra lo mismo, a disfrutar que te lo has ganado.

Tanto, mis compañeros de investigación, como yo te estamos muy agradecidos por el mimo con el que trataste a nuestro trabajo sobre Las Componentes de Influencia. Hemos tenido muy en cuenta tus buenos consejos. Hubiera sido una gran oportunidad, para nosotros, haber podido presentarlo ante tan selecta y entendida audiencia.

Por otra parte, también le hubiera gustado estar ahí a Antonio Sánchez, el cual recorrió contigo las calles de Cádiz en la búsqueda del Vino de Caná. Creo que a su Padre, donde quiera que esté, le ha dado mucha alegría aparecer en tu libro. Para mí fue un descubrimiento saber que eras tan bueno con las letras cómo con los números.

Entrando en otro tema, más mundano, espero que sepas perdonarme que mis dos hijos sean del Barça y no del Real Madrid. En la SEIO de Madrid me dijiste que esto no lo podía consentir. Siguiendo tus consejos les apliqué una transformación ortogonal pero no encontré los valores propios. Mis hijos continúan siendo del Barça.

La Santa Cueva, Haydn y el vino de Jerez están siempre esperando al profesor Jironez que aparezca por Cádiz.

Un fuerte abrazo.

José M^a Gutiérrez (Antonio Sánchez para los amigos).

Index:

A

Albert Satorra, Pompeu Fabra University, Spain. Co-author, Josep M. Oller

CROWD SIZE ESTIMATE BASED ON LINE-DENSITY INTEGRATION AND COMPUTERIZED COUNT OF IMAGES	14
---	-----------

Àlex Sánchez-Pla, University of Barcelona, Spain. Co-authors, Miquel Salicrú, Jordi Ocaña

INTEGRATIVE ANALYSIS OF OMICS DATA USING FUNCTIONAL PROFILES	26
---	-----------

C

Carles M. Cuadras, University of Barcelona, Spain

COMPARING MULTIPLE FACTOR ANALYSIS WITH RELATED METRIC SCALING FOR REPRESENTING MULTIVARIATE DATA.....	5
---	----------

Carles Serrat, Universitat Politècnica de Catalunya – BarcelonaTECH, Spain

JOINT MODELS FOR THE ANALYSIS OF TIME-TO-EVENT DATA WITH LONGITUDINAL INFORMATION	29
--	-----------

Conchita Arenas, University of Barcelona, Spain. Appendix Tribute to Dr. Cuadras

D

Daniel Cuadras, Fundació Sant Joan de Déu, Spain. Appendix Tribute to Dr. Cuadras

Daniel Oberski, Utrecht University, the Netherlands

BAYESIAN MIXTURE PRIORS FOR MEASUREMENT INVARIANCE STRUCTURAL EQUATION MODELING.....	18
---	-----------

Daniel Peña, Carlos III University of Madrid, Spain. Co-author, Andrés M. Alonso

AN UNIFIED PROCEDURE FOR CLUSTERING TIME SERIES	16
--	-----------

E

Elías Moreno, University of Granada, Spain

PROBABILISTIC CLUSTERING: ASYMPTOTIC RESULTS	6
---	----------

Eva Boj and M^a Teresa Costa (University of Barcelona)

MEAN SQUARED ERROR FOR NEW OBSERVATIONS IN LOGISTIC REGRESSION23

Eva Boj, University of Barcelona, Spain

DISTANCE-BASED GENERALIZED LINEAR MODELS: FROM OLS TO IWLS20

G

Germà Coenders, University of Girona, Spain. Co-author, Salvador Linares Mustarós

**LIQUIDITY AND DEBT PROFILES. FROM CLASSIFICATION ACCORDING TO FINANCIAL RATIOS TO
COMPOSITIONAL CLASSIFICATION9**

H

Haroon Idrees, Department of Computer Vision, University of Central Florida, USA

A COMPUTER VISION METHOD FOR IMAGE CROWD COUNTING11

I

Itziar Irigoien, Francesc Mestres and Conchita Arenas (University of Barcelona)

DB-DISCRIMINANT: AN EXTENSION21

J

*Jesús Corral-López (Universitat Politècnica de Catalunya – BarcelonaTECH), Raquel Feixa-Compte
and Maribel Ortego*

**CHANGES IN THE OCCURRENCE OF EXTREME WAVE STORMS IN THE CATALAN COAST
A BAYESIAN APPROACH22**

Joan Manel Batista, ESADE BS. Universitat Ramón Llull, Spain. Appendix Tribute to Dr. Cuadras51

José Fernando Vera, University of Granada, Spain

ON THE ADDED VALUE OF CLUSTERING IN MULTIDIMENSIONAL SCALING17

*José Luis Vicente-Villardón, University of Salamanca, Spain. Co-author, Julio César Hernández-
Sánchez*

LOGISTIC BIPLOTS FOR BINARY, NOMINAL AND ORDINAL DATA19

José M^a Gutiérrez, University of Cádiz, Spain Appendix Tribute to Dr. Cuadras.66

José M^a Gutiérrez, University of Cádiz, Spain. Co-author, Andrés Jiménez

INFLUENCE COMPONENTS25

Josep Fortiana, University of Barcelona, Spain. Appendix Tribute to Dr. Cuadras48

Josep M. Oller, University of Barcelona, Spain. Co-author, Esteban Vegas

ENHANCED PLOTTING IN DATA ANALYSIS.....28

Josep M^a Oller, University of Barcelona, Spain. Appendix Tribute to Dr. Cuadras42

Josep Maria Duran, University of Barcelona, Spain. Appendix Tribute to Dr. Cuadras50

K

Karina Gibert, Universitat Politècnica de Catalunya – BarcelonaTECH, Spain. Appendix Tribute to Dr. Cuadras.....61

Karina Gibert, Universitat Politècnica de Catalunya – BarcelonaTECH, Spain. Co-author, Aida Valls

INTRODUCING SEMANTICS INTO THE DISTANCE-BASED METHODS. AN APPLICATION TO CLUSTERING IN THE FIELD OF TOURISM24

LI

Llorens Badiella, Autonomous University of Barcelona, Spain

ATTENDANCE ESTIMATION FOR MASSIVE DEMONSTRATIONS USING BASIC STATISTICS14

L

Lupe Gómez and Jordi Ocaña, Grup de Recerca en Bioestadística i Bioinformàtica (GRBIO, Grup Consolidat 2014 SGR 464 Generalitat de Catalunya), Spain. Appendix Tribute to Dr. Cuadras31

M

Marc Comas-Cufí, University of Girona, Spain. Co-authors, Josep A. Martín-Fernández, Glòria Mateu-Figueras, Javier Palarea-Albaladejo

A PARAMETRIC APPROACH TO COUNT DATA IN COMPOSITIONAL DATA SETS8

Marcos Cruz, University of Cantabria, Spain. Co-authors, Javier González-Villa, Luis Manuel Cruz-Orive

EFFICIENT AND UNBIASED CROWD SIZE ESTIMATION13

Maribel Ortego, Universitat Politècnica de Catalunya – BarcelonaTECH, Spain. Co-author, Juan José Egozcue

SPURIOUS CORRELATION IN COMPOSITIONAL DATA ANALYSIS: FROM PEARSON TO COPULAS, THROUGH SPEARMAN AND KENDALL6

Martín Ríos, University of Barcelona, Spain. Appendix Tribute to Dr. Cuadras.....46

Mònica Gasulla, Statistical Institute of Catalonia, IDESCAT, Spain. Co-authors, Jesús Muñoz, Jonathan Jorba, Sergio Plaza

DYNAMIC FACTOR MODELS FOR GDP NOWCASTING: AN APPLICATION TO THE CATALAN ECONOMY.....25

P

Purificación Galindo, University of Salamanca, Spain. Appendix Tribute to Dr. Cuadras.....53

List of participants:

Name	Affiliation	e-mail
Yovaninna Alarcón	<i>(Universitat Politècnica de Catalunya - BarcelonaTECH)</i>	yovi.alarcon@gmail.com
Juan Amor Ferrando	<i>(Digital Transformation Services Consultant)</i>	juanamor99@gmail.com
Concepción Arenas Sola	<i>(Universitat de Barcelona)</i>	carenas@ub.edu
Llorens Badiella Busquets	<i>(Universitat Autònoma de Barcelona)</i>	llorenc.badiella@uab.cat
Joan Manel Batista Foguet	<i>(ESADE BS. Universitat Ramón Llull.)</i>	joanm.batista@esade.edu
Marta Bofill	<i>(Universitat Politècnica de Catalunya - BarcelonaTECH)</i>	marta.bofill.roig@upc.edu
Eva Boj del Val	<i>(Universitat de Barcelona)</i>	evaboj@ub.edu
Francesc Carmona Pontaque	<i>(Universitat de Barcelona)</i>	fcarmona@ub.edu
Marta Casals	<i>(Institut d'Estadística de Catalunya)</i>	mcasals@idescat.cat
Germà Coenders Gallart	<i>(Universitat de Girona)</i>	germa.coenders@udg.edu
Marc Comas Cufí	<i>(Universitat de Girona)</i>	mcomas@imae.udg.edu
Jesús Corral López	<i>(Universitat Politècnica de Catalunya - BarcelonaTECH)</i>	jesus.corral@upc.edu
M ^a Teresa Costa Cor	<i>(Universitat de Barcelona)</i>	tcosta@ub.edu
Carles M ^a Cuadras Avellana	<i>(Universitat de Barcelona)</i>	ccuadras@ub.edu
Daniel Cuadras Pallejà	<i>(Fundació Sant Joan de Déu)</i>	d.cuadras@pssjd.org
Marcos Cruz Rodríguez	<i>(Universidad de Cantabria)</i>	marcos.cruz@unican.es
Luis Manuel Cruz Orive	<i>(Universidad de Cantabria)</i>	luis.cruz@unican.es
Pedro Delicado Users	<i>(Universitat Politècnica de Catalunya - BarcelonaTECH)</i>	pedro.delicado@upc.edu
Josep Maria Duran Rubies	<i>(Universitat de Barcelona)</i>	jduran16@xtec.cat
Juan José Egozcue	<i>(Universitat Politècnica de Catalunya - BarcelonaTECH)</i>	juan.jose.egozcue@upc.edu
Anna Esteve Gómez	<i>(Hospital Universitari Germans Trias i Pujol)</i>	aesteve@iconcologia.net
Raquel Feixa Compte	<i>(Universitat Politècnica de Catalunya - BarcelonaTECH)</i>	raquel.feixa.compte@gmail.com
Josep Fortiana Gregori	<i>(Universitat de Barcelona)</i>	fortiana@ub.edu
Purificación Galindo Villardón	<i>(Universidad de Salamanca)</i>	pgalindo@usal.es
Mònica Gasulla i Ramon	<i>(Institut d'Estadística de Catalunya)</i>	mgasulla@idescat.cat
Karina Gibert	<i>(Universitat Politècnica de Catalunya - BarcelonaTECH)</i>	karina.gibert@upc.edu

Lupe Gómez Melis	<i>(Universitat Politècnica de Catalunya - BarcelonaTECH)</i>	lupe.gomez@upc.edu
Moisés Gómez Mateu	<i>(Universitat Politècnica de Catalunya - BarcelonaTECH)</i>	moises.gomez.mateu@upc.edu
Javier González Villa	<i>(Universidad de Cantabria)</i>	gvillaj@unican.es
Joan Guàrdia Olmos	<i>(Societat Catalana d'Estadística)</i>	jguardia@ub.edu
José M ^a Gutiérrez Pérez	<i>(Universidad de Cádiz)</i>	josema.gutierrez@uca.es
Julio César Hernández Sánchez	<i>(Instituto Nacional de Estadística)</i>	juliocesar.hernandez.sanchez@ine.es
Haroon Idrees	<i>(University of Central Florida)</i>	haroonidrees@gmail.com
Itziar Irigoien Garbizu	<i>(Universidad del País Vasco)</i>	itziar.irigoien@ehu.es
Eusebio Jarauta Bragulat	<i>(Universitat Politècnica de Catalunya - BarcelonaTECH)</i>	eusebi.jarauta@upc.edu
Andrés Jiménez Jiménez	<i>(Universidad de Cádiz)</i>	andres.jimenez@uca.es
Jonatan Jorba Giménez	<i>(Institut d'Estadística de Catalunya)</i>	jjorba@idescat.cat
Salvador Linares Mustarós	<i>(Universitat de Girona)</i>	salvador.linares@udg.edu
Felipe de Mendiburu Delgado	<i>(Universitat Politècnica de Catalunya - BarcelonaTECH)</i>	fdemendiburu@yahoo.com
Josep Antoni Martín Fernández	<i>(Universitat de Girona)</i>	josepantoni.martin@udg.edu
Elías Moreno Bas	<i>(Universidad de Granada)</i>	emoreno@ugr.es
Daniel Oberski	<i>(Utrecht University)</i>	d.l.oberski@uu.nl
Josep M ^a Oller Sala	<i>(Universitat de Barcelona)</i>	joller@ub.edu
Jordi Ocaña Rebull	<i>(Universitat de Barcelona)</i>	jocana@ub.edu
Glòria Mateu Figueras	<i>(Universitat de Girona)</i>	gloria.mateu@udg.edu
Francesc Mestres Naval	<i>(Universitat de Barcelona)</i>	fmestres@ub.edu
Jesús Muñoz Malo	<i>(Institut d'Estadística de Catalunya)</i>	mailto:jmalo@idescat.cat
Maribel Ortego Martínez	<i>(Universitat Politècnica de Catalunya - BarcelonaTECH)</i>	ma.isabel.ortego@upc.edu
Javier Palarea Albaladejo	<i>(Biomathematics and Statistics Scotland)</i>	javier.palarea@bioss.ac.uk
Daniel Peña Sánchez de Rivera	<i>(Universidad Carlos III de Madrid)</i>	daniel.pena@uc3m.es
Sergio Plaza Palacios	<i>(Institut d'Estadística de Catalunya)</i>	splaza@idescat.cat
Martín Ríos Alcolea	<i>(Universitat de Barcelona)</i>	mrios@ub.edu
Cristina Rovira Trepal	<i>(Institut d'Estadística de Catalunya)</i>	crovira@idescat.cat
Miquel Salicrú Pagés	<i>(Universitat de Barcelona)</i>	msalicru@ub.edu
Àlex Sánchez Pla	<i>(Universitat de Barcelona)</i>	asanchez@ub.edu
Albert Satorra Brucart	<i>(Universitat Pompeu Fabra)</i>	albert.satorra@upf.edu

Carles Serrat i Pie	<i>(Universitat Politècnica de Catalunya - BarcelonaTECH)</i>	carles.serrat@upc.edu
Frederic Udina i Abelló	<i>(Institut d'Estadística de Catalunya)</i>	fudina@idescat.cat
Aida Valls Mateu	<i>(Universitat Rovira i Virgili)</i>	aida.valls@urv.net
José Fernando Vera	<i>(Universidad de Granada)</i>	jfvera@ugr.es
Esteban Vegas Lozano	<i>(Universitat de Barcelona)</i>	evegas@ub.edu
José Luis Vicente Villardón	<i>(Universidad de Salamanca)</i>	villardon@usal.es



In visual order:

José Fernando Vera (University of Granada)

Carles M. Cuadras (University of Barcelona)

Eva Boj (University of Barcelona)



<http://www.ub.edu/wamyc/>